airline-lor-project

October 8, 2023

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

[2]:
```
df=pd.read_excel(r"C:
↪\Users\prera\OneDrive\Desktop\Imarticus\ML\datasets\airline data.xlsx")
```

[3]:
```
df
```

[3]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	\
0	0	19556	Female	Loyal Customer	52	Business travel	
1	1	90035	Female	Loyal Customer	36	Business travel	
2	2	12360	Male	disloyal Customer	20	Business travel	
3	3	77959	Male	Loyal Customer	44	Business travel	
4	4	36875	Female	Loyal Customer	49	Business travel	
…	…	…	…	…	…	…	
25971	25971	78463	Male	disloyal Customer	34	Business travel	
25972	25972	71167	Male	Loyal Customer	23	Business travel	
25973	25973	37675	Female	Loyal Customer	17	Personal Travel	
25974	25974	90086	Male	Loyal Customer	14	Business travel	
25975	25975	34799	Female	Loyal Customer	42	Personal Travel	

	Class	Flight Distance	Inflight wifi service	\
0	Eco	160.0	5.0	
1	Business	2863.0	1.0	
2	Eco	192.0	2.0	
3	Business	3377.0	0.0	
4	Eco	1182.0	2.0	
…	…	…	…	
25971	Business	526.0	3.0	
25972	Business	646.0	4.0	
25973	Eco	828.0	2.0	
25974	Business	1127.0	3.0	
25975	Eco	264.0	2.0	

```
       Departure/Arrival time convenient  …  Inflight entertainment  \
0                                    4.0  …                     5.0
1                                    1.0  …                     4.0
2                                    0.0  …                     2.0
3                                    0.0  …                     1.0
4                                    3.0  …                     2.0
…                                    …  …                       …
25971                                3.0  …                     4.0
25972                                4.0  …                     4.0
25973                                5.0  …                     2.0
25974                                3.0  …                     4.0
25975                                5.0  …                     1.0

       On-board service  Leg room service  Baggage handling  Checkin service  \
0                   5.0               5.0               5.0              2.0
1                   4.0               4.0               4.0              3.0
2                   4.0               1.0               3.0              2.0
3                   1.0               1.0               1.0              3.0
4                   2.0               2.0               2.0              4.0
…                   …                 …                 …                …
25971               3.0               2.0               4.0              4.0
25972               4.0               5.0               5.0              5.0
25973               4.0               3.0               4.0              5.0
25974               3.0               2.0               5.0              4.0
25975               1.0               2.0               1.0              1.0

       Inflight service  Cleanliness  Departure Delay in Minutes  \
0                   5.0          5.0                        50.0
1                   4.0          5.0                         0.0
2                   2.0          2.0                         0.0
3                   1.0          4.0                         0.0
4                   2.0          4.0                         0.0
…                   …            …                           …
25971               5.0          4.0                         0.0
25972               5.0          4.0                         0.0
25973               4.0          2.0                         0.0
25974               5.0          4.0                         0.0
25975               1.0          1.0                         0.0

       Arrival Delay in Minutes  satisfaction
0                          44.0     satisfied
1                           0.0     satisfied
2                           0.0  dissatisfied
3                           6.0     satisfied
4                          20.0     satisfied
…                          …           …
25971                       0.0  dissatisfied
```

```
25972                       0.0     satisfied
25973                       0.0  dissatisfied
25974                       0.0     satisfied
25975                       0.0  dissatisfied

[25976 rows x 25 columns]
```

[4]:
```
df1=df.copy(deep=True)
df2=df.copy(deep=True)
```

[5]:
```
df.shape
```

[5]:
```
(25976, 25)
```

[6]:
```
df.head()
```

[6]:
```
   Unnamed: 0     id  Gender      Customer Type  Age   Type of Travel  \
0           0  19556  Female     Loyal Customer   52  Business travel
1           1  90035  Female     Loyal Customer   36  Business travel
2           2  12360    Male  disloyal Customer   20  Business travel
3           3  77959    Male     Loyal Customer   44  Business travel
4           4  36875  Female     Loyal Customer   49  Business travel

      Class  Flight Distance  Inflight wifi service  \
0       Eco            160.0                    5.0
1  Business           2863.0                    1.0
2       Eco            192.0                    2.0
3  Business           3377.0                    0.0
4       Eco           1182.0                    2.0

   Departure/Arrival time convenient  …  Inflight entertainment  \
0                                4.0  …                    5.0
1                                1.0  …                    4.0
2                                0.0  …                    2.0
3                                0.0  …                    1.0
4                                3.0  …                    2.0

   On-board service  Leg room service  Baggage handling  Checkin service  \
0               5.0               5.0               5.0              2.0
1               4.0               4.0               4.0              3.0
2               4.0               1.0               3.0              2.0
3               1.0               1.0               1.0              3.0
4               2.0               2.0               2.0              4.0

   Inflight service  Cleanliness  Departure Delay in Minutes  \
0               5.0          5.0                        50.0
1               4.0          5.0                         0.0
```

```
2               2.0          2.0                          0.0
3               1.0          4.0                          0.0
4               2.0          4.0                          0.0

   Arrival Delay in Minutes  satisfaction
0                      44.0     satisfied
1                       0.0     satisfied
2                       0.0  dissatisfied
3                       6.0     satisfied
4                      20.0     satisfied

[5 rows x 25 columns]
```

[7]:
```
df.tail()
```

[7]:
```
       Unnamed: 0     id  Gender      Customer Type  Age   Type of Travel  \
25971       25971  78463    Male  disloyal Customer   34  Business travel
25972       25972  71167    Male     Loyal Customer   23  Business travel
25973       25973  37675  Female     Loyal Customer   17  Personal Travel
25974       25974  90086    Male     Loyal Customer   14  Business travel
25975       25975  34799  Female     Loyal Customer   42  Personal Travel

          Class  Flight Distance  Inflight wifi service  \
25971  Business            526.0                    3.0
25972  Business            646.0                    4.0
25973       Eco            828.0                    2.0
25974  Business           1127.0                    3.0
25975       Eco            264.0                    2.0

       Departure/Arrival time convenient  …  Inflight entertainment  \
25971                                3.0  …                     4.0
25972                                4.0  …                     4.0
25973                                5.0  …                     2.0
25974                                3.0  …                     4.0
25975                                5.0  …                     1.0

       On-board service  Leg room service  Baggage handling  Checkin service  \
25971               3.0               2.0               4.0              4.0
25972               4.0               5.0               5.0              5.0
25973               4.0               3.0               4.0              5.0
25974               3.0               2.0               5.0              4.0
25975               1.0               2.0               1.0              1.0

       Inflight service  Cleanliness  Departure Delay in Minutes  \
25971               5.0          4.0                         0.0
25972               5.0          4.0                         0.0
25973               4.0          2.0                         0.0
```

```
25974               5.0          4.0                          0.0
25975               1.0          1.0                          0.0

       Arrival Delay in Minutes  satisfaction
25971                       0.0  dissatisfied
25972                       0.0     satisfied
25973                       0.0  dissatisfied
25974                       0.0     satisfied
25975                       0.0  dissatisfied

[5 rows x 25 columns]
```

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25976 entries, 0 to 25975
Data columns (total 25 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Unnamed: 0                         25976 non-null  int64
 1   id                                 25976 non-null  int64
 2   Gender                             25976 non-null  object
 3   Customer Type                      25976 non-null  object
 4   Age                                25976 non-null  int64
 5   Type of Travel                     25976 non-null  object
 6   Class                              25976 non-null  object
 7   Flight Distance                    25751 non-null  float64
 8   Inflight wifi service              25935 non-null  float64
 9   Departure/Arrival time convenient  25779 non-null  float64
 10  Ease of Online booking             25728 non-null  float64
 11  Gate location                      25766 non-null  float64
 12  Food and drink                     25846 non-null  float64
 13  Online boarding                    25891 non-null  float64
 14  Seat comfort                       25866 non-null  float64
 15  Inflight entertainment             25845 non-null  float64
 16  On-board service                   25841 non-null  float64
 17  Leg room service                   25862 non-null  float64
 18  Baggage handling                   25830 non-null  float64
 19  Checkin service                    25818 non-null  float64
 20  Inflight service                   25779 non-null  float64
 21  Cleanliness                        25830 non-null  float64
 22  Departure Delay in Minutes         25899 non-null  float64
 23  Arrival Delay in Minutes           25883 non-null  float64
 24  satisfaction                       25976 non-null  object
dtypes: float64(17), int64(3), object(5)
memory usage: 5.0+ MB
```

[9]:
```
df.describe()
```

[9]:

	Unnamed: 0	id	Age	Flight Distance
count	25976.000000	25976.000000	25976.000000	25751.000000
mean	12987.500000	65005.657992	39.620958	1193.324919
std	7498.769632	37611.526647	15.135685	998.547425
min	0.000000	17.000000	7.000000	31.000000
25%	6493.750000	32170.500000	27.000000	414.000000
50%	12987.500000	65319.500000	40.000000	848.000000
75%	19481.250000	97584.250000	51.000000	1744.000000
max	25975.000000	129877.000000	85.000000	4983.000000

	Inflight wifi service	Departure/Arrival time convenient
count	25935.000000	25779.000000
mean	2.724619	3.045657
std	1.335419	1.533312
min	0.000000	0.000000
25%	2.000000	2.000000
50%	3.000000	3.000000
75%	4.000000	4.000000
max	5.000000	5.000000

	Ease of Online booking	Gate location	Food and drink	Online boarding
count	25728.000000	25766.000000	25846.000000	25891.000000
mean	2.755403	2.976908	3.214308	3.261365
std	1.412670	1.281423	1.331842	1.355765
min	0.000000	1.000000	0.000000	0.000000
25%	2.000000	2.000000	2.000000	2.000000
50%	3.000000	3.000000	3.000000	4.000000
75%	4.000000	4.000000	4.000000	4.000000
max	5.000000	5.000000	5.000000	5.000000

	Seat comfort	Inflight entertainment	On-board service
count	25866.000000	25845.000000	25841.000000
mean	3.449818	3.358212	3.385782
std	1.320003	1.337980	1.282007
min	1.000000	0.000000	0.000000
25%	2.000000	2.000000	2.000000
50%	4.000000	4.000000	4.000000
75%	5.000000	4.000000	4.000000
max	5.000000	5.000000	5.000000

	Leg room service	Baggage handling	Checkin service	Inflight service
count	25862.000000	25830.000000	25818.000000	25779.000000
mean	3.350012	3.632753	3.313967	3.648629
std	1.319025	1.176418	1.269188	1.181216
min	0.000000	1.000000	1.000000	0.000000

```
25%            2.000000          3.000000          3.000000          3.000000
50%            4.000000          4.000000          3.000000          4.000000
75%            4.000000          5.000000          4.000000          5.000000
max            5.000000          5.000000          5.000000          5.000000

        Cleanliness  Departure Delay in Minutes  Arrival Delay in Minutes
count  25830.000000                25899.000000              25883.000000
mean       3.286527                   14.300398                 14.745084
std        1.319237                   37.425147                 37.523901
min        0.000000                    0.000000                  0.000000
25%        2.000000                    0.000000                  0.000000
50%        3.000000                    0.000000                  0.000000
75%        4.000000                   12.000000                 13.000000
max        5.000000                 1128.000000               1115.000000
```

[10]:
```
df.nunique()
```

[10]:
```
Unnamed: 0                          25976
id                                  25976
Gender                                  2
Customer Type                           2
Age                                    75
Type of Travel                          2
Class                                   3
Flight Distance                      3272
Inflight wifi service                   6
Departure/Arrival time convenient       6
Ease of Online booking                  6
Gate location                           5
Food and drink                          6
Online boarding                         6
Seat comfort                            5
Inflight entertainment                  6
On-board service                        6
Leg room service                        6
Baggage handling                        5
Checkin service                         5
Inflight service                        6
Cleanliness                             6
Departure Delay in Minutes            313
Arrival Delay in Minutes              320
satisfaction                            2
dtype: int64
```

[11]:
```
df.duplicated(subset=None, keep='first')
```

[11]:
```
0        False
1        False
2        False
3        False
4        False
         …
25971    False
25972    False
25973    False
25974    False
25975    False
Length: 25976, dtype: bool
```

[12]:
```
dF=df.drop_duplicates()
```

[13]:
```
dF.shape #there are no duplicate values in out dataset
```

[13]:
```
(25976, 25)
```

[14]:
```
df.isnull().sum()
```

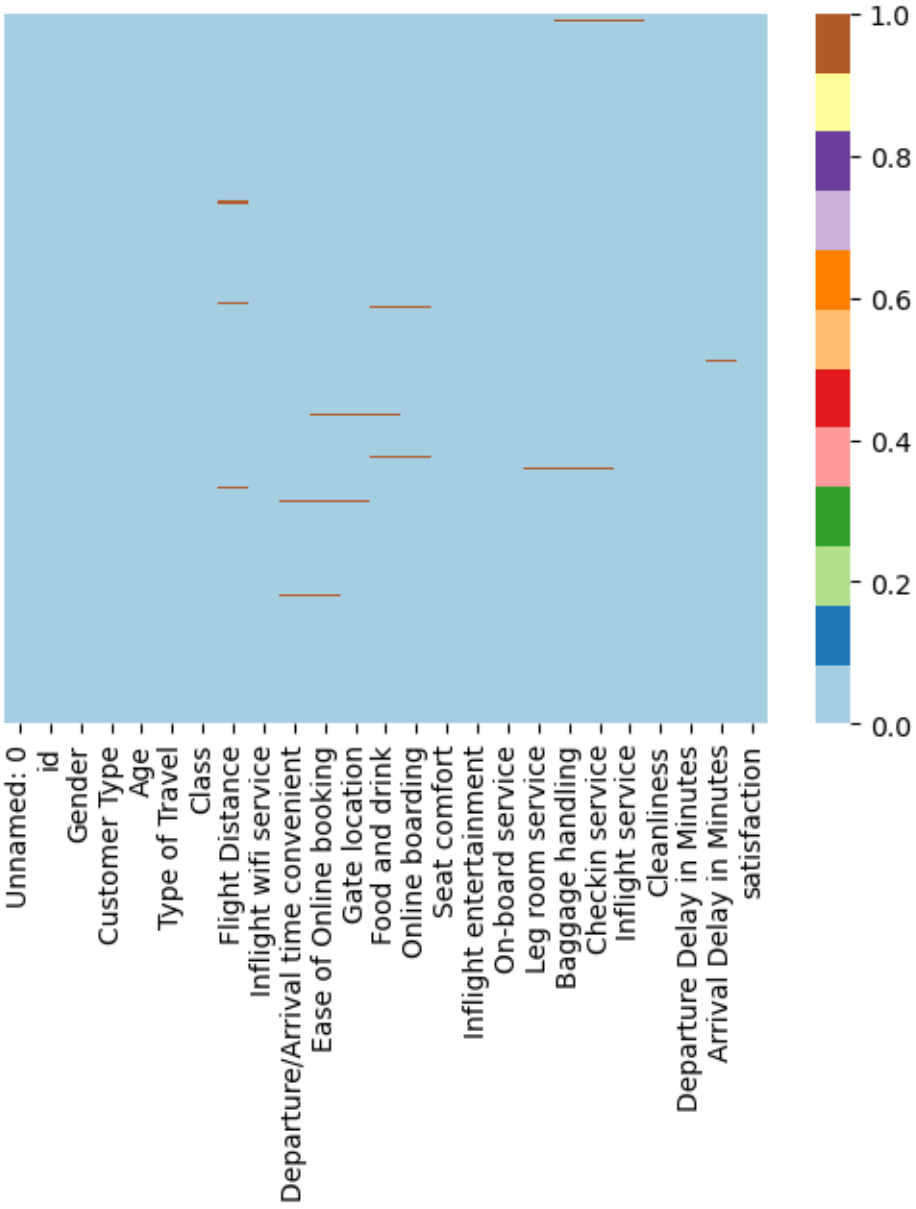
[14]:
```
Unnamed: 0                           0
id                                   0
Gender                               0
Customer Type                        0
Age                                  0
Type of Travel                       0
Class                                0
Flight Distance                    225
Inflight wifi service               41
Departure/Arrival time convenient  197
Ease of Online booking             248
Gate location                      210
Food and drink                     130
Online boarding                     85
Seat comfort                       110
Inflight entertainment             131
On-board service                   135
Leg room service                   114
Baggage handling                   146
Checkin service                    158
Inflight service                   197
Cleanliness                        146
Departure Delay in Minutes          77
Arrival Delay in Minutes            93
satisfaction                         0
dtype: int64
```

[15]:
```
(df.isnull().sum()/len(df))*100
```

[15]:
```
Unnamed: 0                          0.000000
id                                  0.000000
Gender                              0.000000
Customer Type                       0.000000
Age                                 0.000000
Type of Travel                      0.000000
Class                               0.000000
Flight Distance                     0.866184
Inflight wifi service               0.157838
Departure/Arrival time convenient   0.758392
Ease of Online booking              0.954727
Gate location                       0.808439
Food and drink                      0.500462
Online boarding                     0.327225
Seat comfort                        0.423468
Inflight entertainment              0.504312
On-board service                    0.519711
Leg room service                    0.438867
Baggage handling                    0.562057
Checkin service                     0.608254
Inflight service                    0.758392
Cleanliness                         0.562057
Departure Delay in Minutes          0.296427
Arrival Delay in Minutes            0.358023
satisfaction                        0.000000
dtype: float64
```

[16]:
```
sns.heatmap(df.isnull(),yticklabels=False, cmap='Paired')
```

[16]:
```
<Axes: >
```



[17]:
```
df.rename(columns={'Unnamed: 0':'sr_no','Departure/Arrival time convenient':
 ↪'Departure_Arrival time convenient',
                   'On-board service':'Onboard service'},inplace=True)
```

[18]:
```
df.columns
```

[18]:
```
Index(['sr_no', 'id', 'Gender', 'Customer Type', 'Age', 'Type of Travel',
       'Class', 'Flight Distance', 'Inflight wifi service',
       'Departure_Arrival time convenient', 'Ease of Online booking',
       'Gate location', 'Food and drink', 'Online boarding', 'Seat comfort',
```

```
       'Inflight entertainment', 'Onboard service', 'Leg room service',
       'Baggage handling', 'Checkin service', 'Inflight service',
       'Cleanliness', 'Departure Delay in Minutes', 'Arrival Delay in Minutes',
       'satisfaction'],
      dtype='object')
```

[19]:
```
#dropping all the rows with less than 100 missing values
df.dropna(subset=['Inflight wifi service'],axis=0,inplace=True)
df.dropna(subset=['Online boarding'],axis=0,inplace=True)
df.dropna(subset=['Departure Delay in Minutes'],axis=0,inplace=True)
df.dropna(subset=['Arrival Delay in Minutes'],axis=0,inplace=True)
```

[20]:
```
df['Flight Distance'].fillna(df['Flight Distance'].median(),inplace=True)
df['Departure_Arrival time convenient'].ffill(axis=0,inplace=True)
df['Ease of Online booking'].bfill(axis=0,inplace=True)
df['Gate location'].ffill(axis=0,inplace=True)
df['Food and drink'].bfill(axis=0,inplace=True)
df['Seat comfort'].ffill(axis=0,inplace=True)
df['Inflight entertainment'].bfill(axis=0,inplace=True)
df['Onboard service'].ffill(axis=0,inplace=True)
df['Leg room service'].bfill(axis=0,inplace=True)
df['Baggage handling'].ffill(axis=0,inplace=True)
df['Checkin service'].bfill(axis=0,inplace=True)
df['Inflight service'].ffill(axis=0,inplace=True)
df['Cleanliness'].bfill(axis=0,inplace=True)
```

[21]:
```
df.isnull().sum()
```

[21]:
```
sr_no                              0
id                                 0
Gender                             0
Customer Type                      0
Age                                0
Type of Travel                     0
Class                              0
Flight Distance                    0
Inflight wifi service              0
Departure_Arrival time convenient  0
Ease of Online booking             0
Gate location                      0
Food and drink                     0
Online boarding                    0
Seat comfort                       0
Inflight entertainment             0
Onboard service                    0
Leg room service                   0
Baggage handling                   0
```

```
Checkin service                    0
Inflight service                   0
Cleanliness                        0
Departure Delay in Minutes         0
Arrival Delay in Minutes           0
satisfaction                       0
dtype: int64
```

[22]:
```
df['Gender'].value_counts()
```

[22]:
```
Female    13011
Male      12680
Name: Gender, dtype: int64
```

[23]:
```
df['Customer Type'].value_counts()
```

[23]:
```
Loyal Customer       20945
disloyal Customer     4746
Name: Customer Type, dtype: int64
```

[24]:
```
df['Type of Travel'].value_counts()
```

[24]:
```
Business travel    17838
Personal Travel     7853
Name: Type of Travel, dtype: int64
```

[25]:
```
df['Class'].value_counts()
```

[25]:
```
Business    12362
Eco         11428
Eco Plus     1901
Name: Class, dtype: int64
```

[26]:
```
cat_data=df.select_dtypes(include=object)
num_data=df.select_dtypes(exclude=object)
```

[27]:
```
cat_data
```

[27]:

	Gender	Customer Type	Type of Travel	Class	satisfaction
0	Female	Loyal Customer	Business travel	Eco	satisfied
1	Female	Loyal Customer	Business travel	Business	satisfied
2	Male	disloyal Customer	Business travel	Eco	dissatisfied
3	Male	Loyal Customer	Business travel	Business	satisfied
4	Female	Loyal Customer	Business travel	Eco	satisfied
…	…	…	…	…	…
25971	Male	disloyal Customer	Business travel	Business	dissatisfied
25972	Male	Loyal Customer	Business travel	Business	satisfied

```
25973  Female     Loyal Customer  Personal Travel       Eco  dissatisfied
25974    Male     Loyal Customer  Business travel  Business     satisfied
25975  Female     Loyal Customer  Personal Travel       Eco  dissatisfied

[25691 rows x 5 columns]
```

[28]:
```
num_data
```

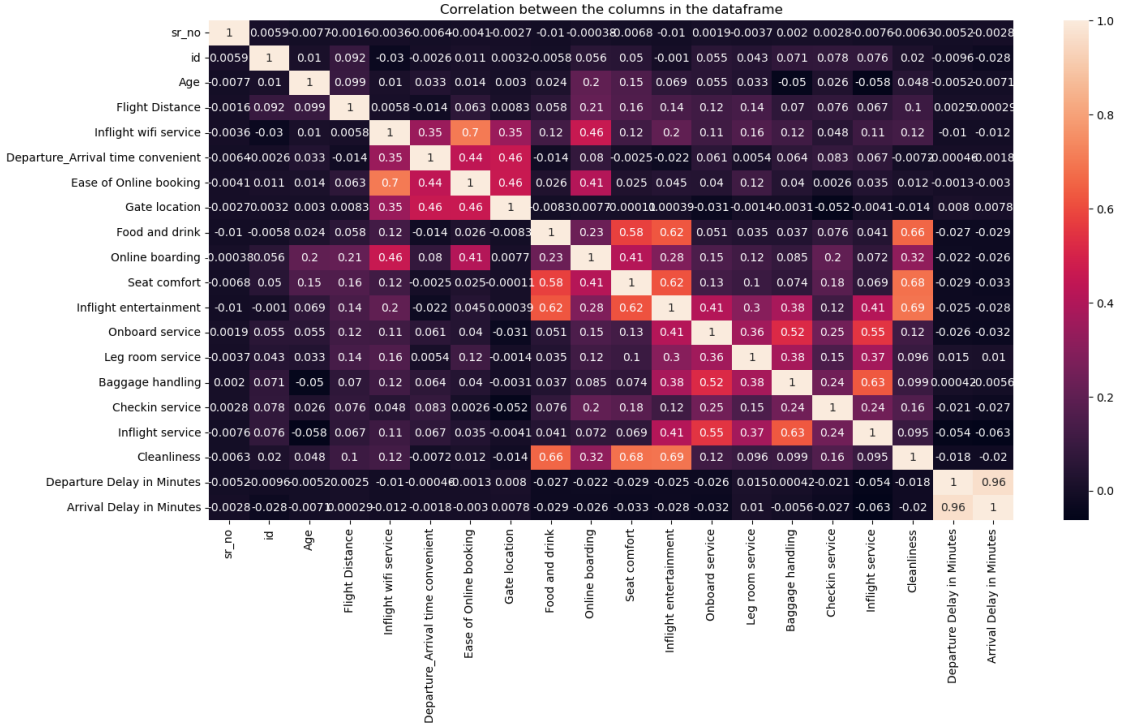
[28]:
```
       sr_no     id  Age  Flight Distance  Inflight wifi service  \
0          0  19556   52            160.0                    5.0
1          1  90035   36           2863.0                    1.0
2          2  12360   20            192.0                    2.0
3          3  77959   44           3377.0                    0.0
4          4  36875   49           1182.0                    2.0
…        …    …  …               …                        …
25971  25971  78463   34            526.0                    3.0
25972  25972  71167   23            646.0                    4.0
25973  25973  37675   17            828.0                    2.0
25974  25974  90086   14           1127.0                    3.0
25975  25975  34799   42            264.0                    2.0

       Departure_Arrival time convenient  Ease of Online booking  \
0                                    4.0                     3.0
1                                    1.0                     3.0
2                                    0.0                     2.0
3                                    0.0                     0.0
4                                    3.0                     4.0
…                                    …                       …
25971                                3.0                     3.0
25972                                4.0                     4.0
25973                                5.0                     1.0
25974                                3.0                     3.0
25975                                5.0                     2.0

       Gate location  Food and drink  Online boarding  Seat comfort  \
0                4.0             3.0              4.0           3.0
1                1.0             5.0              4.0           5.0
2                4.0             2.0              2.0           2.0
3                2.0             3.0              4.0           4.0
4                3.0             4.0              1.0           2.0
…              …               …                …             …
25971            1.0             4.0              3.0           4.0
25972            4.0             4.0              4.0           4.0
25973            5.0             2.0              1.0           2.0
25974            3.0             4.0              4.0           4.0
25975            5.0             4.0              2.0           2.0
```

```
       Inflight entertainment  Onboard service  Leg room service  \
0                         5.0              5.0               5.0
1                         4.0              4.0               4.0
2                         2.0              4.0               1.0
3                         1.0              1.0               1.0
4                         2.0              2.0               2.0
…                         …                …                 …
25971                     4.0              3.0               2.0
25972                     4.0              4.0               5.0
25973                     2.0              4.0               3.0
25974                     4.0              3.0               2.0
25975                     1.0              1.0               2.0

       Baggage handling  Checkin service  Inflight service  Cleanliness  \
0                   5.0              2.0               5.0          5.0
1                   4.0              3.0               4.0          5.0
2                   3.0              2.0               2.0          2.0
3                   1.0              3.0               1.0          4.0
4                   2.0              4.0               2.0          4.0
…                   …                …                 …            …
25971               4.0              4.0               5.0          4.0
25972               5.0              5.0               5.0          4.0
25973               4.0              5.0               4.0          2.0
25974               5.0              4.0               5.0          4.0
25975               1.0              1.0               1.0          1.0

       Departure Delay in Minutes  Arrival Delay in Minutes
0                            50.0                      44.0
1                             0.0                       0.0
2                             0.0                       0.0
3                             0.0                       6.0
4                             0.0                      20.0
…                             …                         …
25971                         0.0                       0.0
25972                         0.0                       0.0
25973                         0.0                       0.0
25974                         0.0                       0.0
25975                         0.0                       0.0

[25691 rows x 20 columns]
```

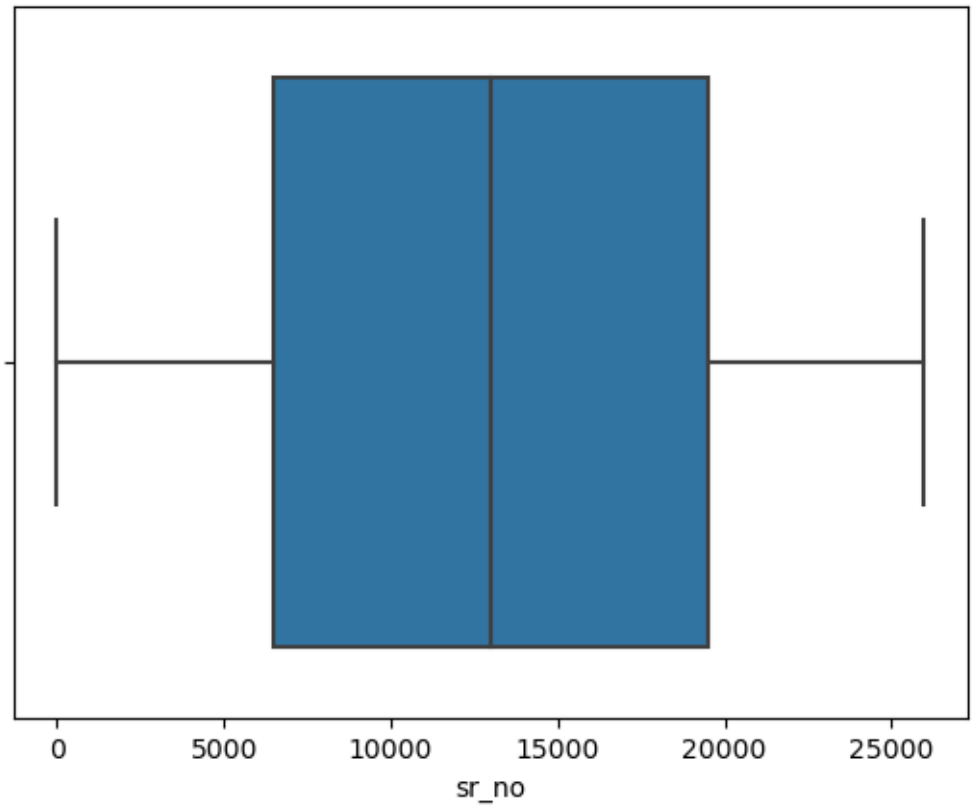
[29]:
```
plt.figure(figsize=(16,8))
sns.heatmap(df.corr(),annot=True)
plt.title('Correlation between the columns in the dataframe')
plt.show()
```

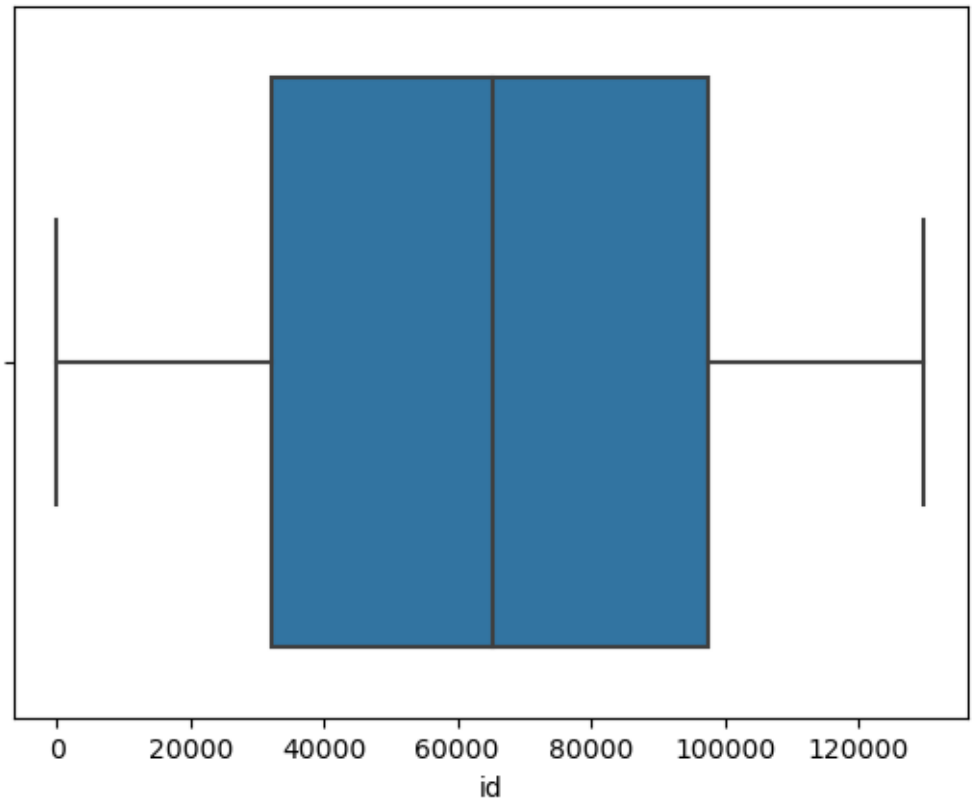
```
C:\Users\prera\AppData\Local\Temp\ipykernel_12352\287062504.py:2: FutureWarning:
```

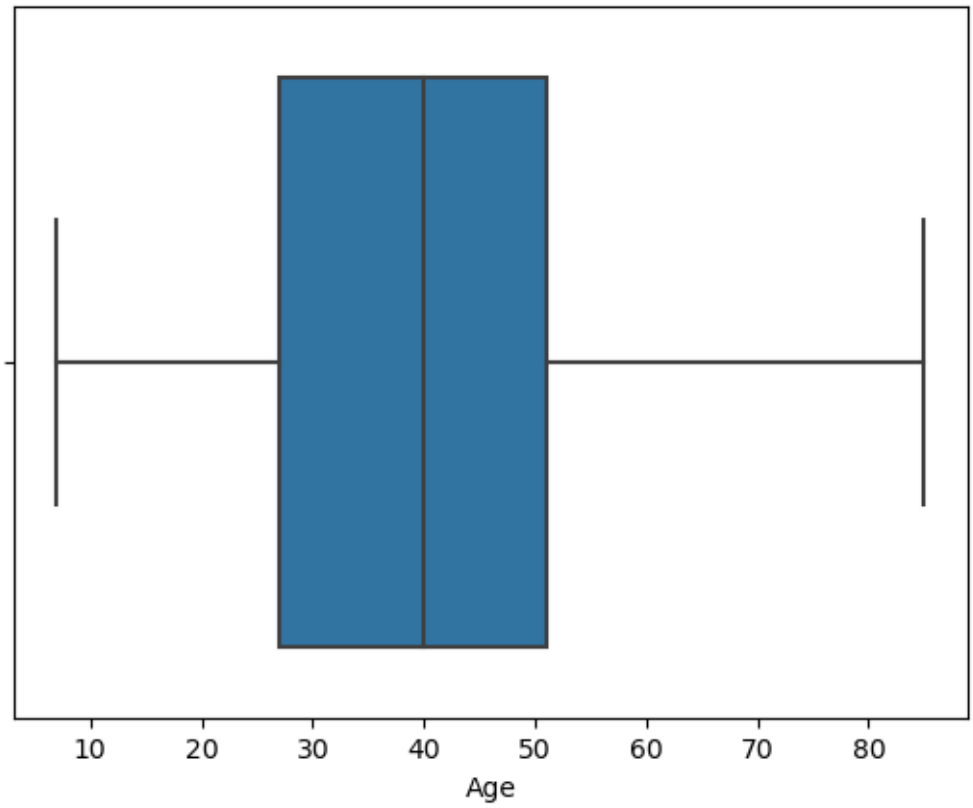
```
The default value of numeric_only in DataFrame.corr is deprecated. In a future
version, it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  sns.heatmap(df.corr(),annot=True)
```

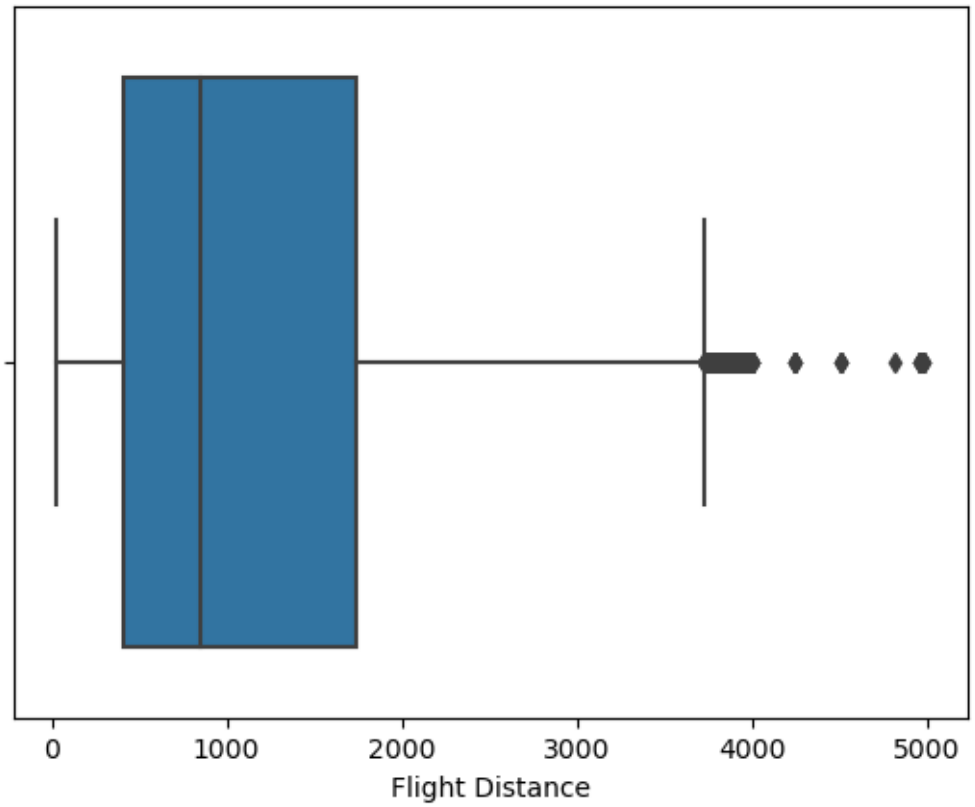


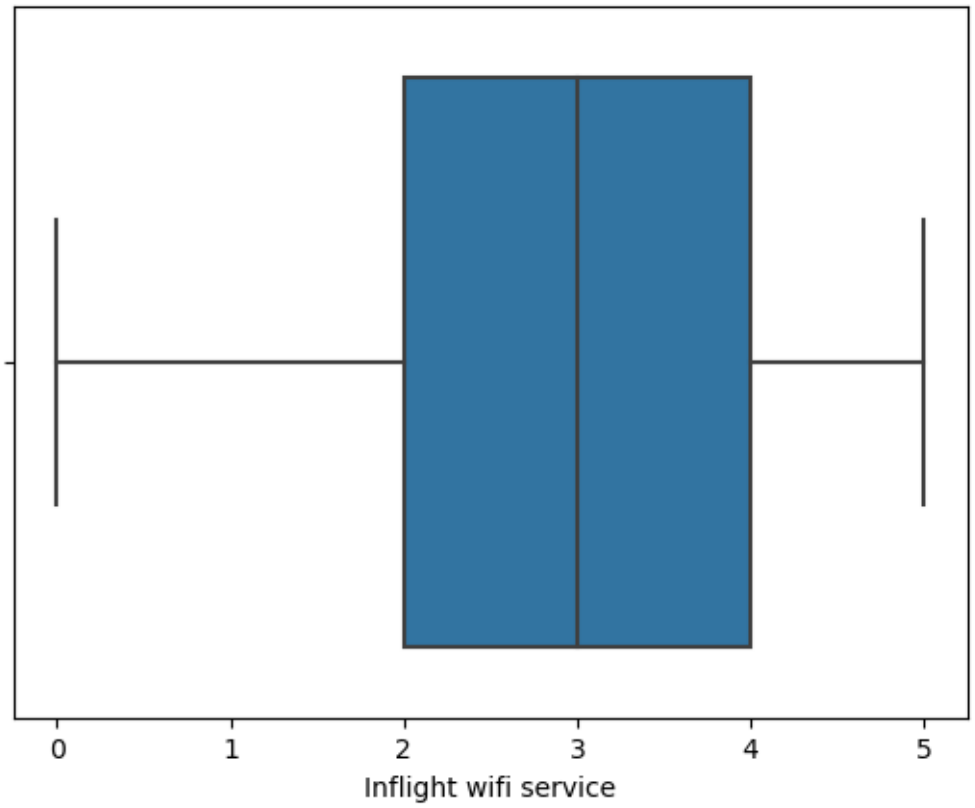
[30]:
```
for i in num_data:
    sns.boxplot(x=df[i])
    plt.show()
```

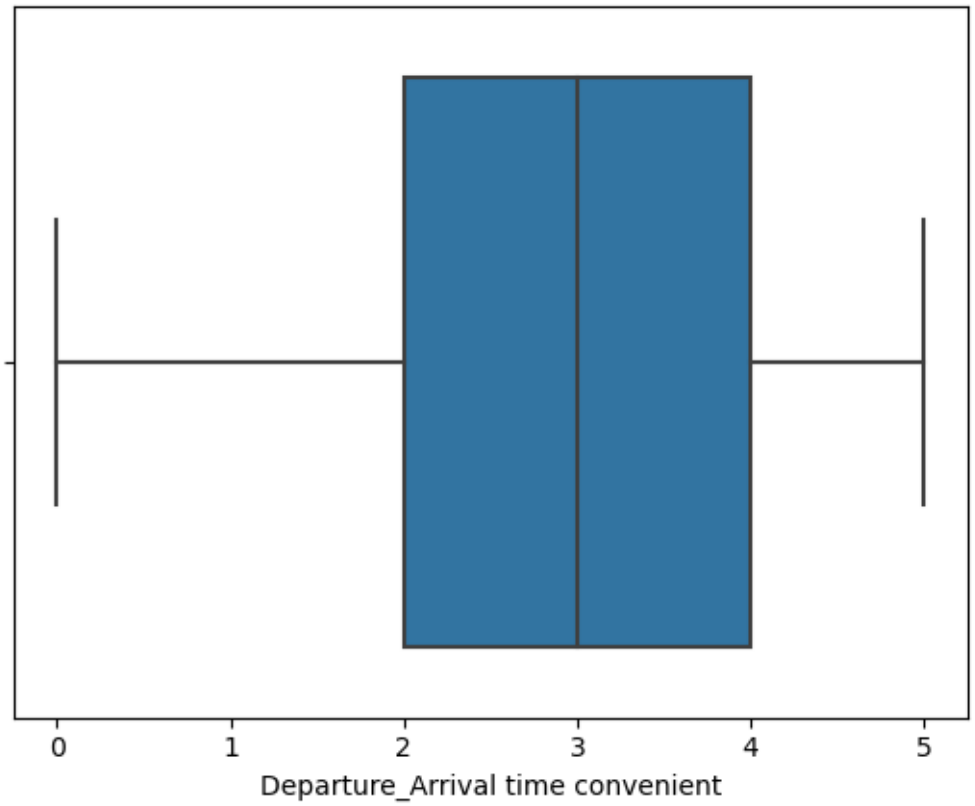

0
5000
10000
15000
20000
25000
sr_no

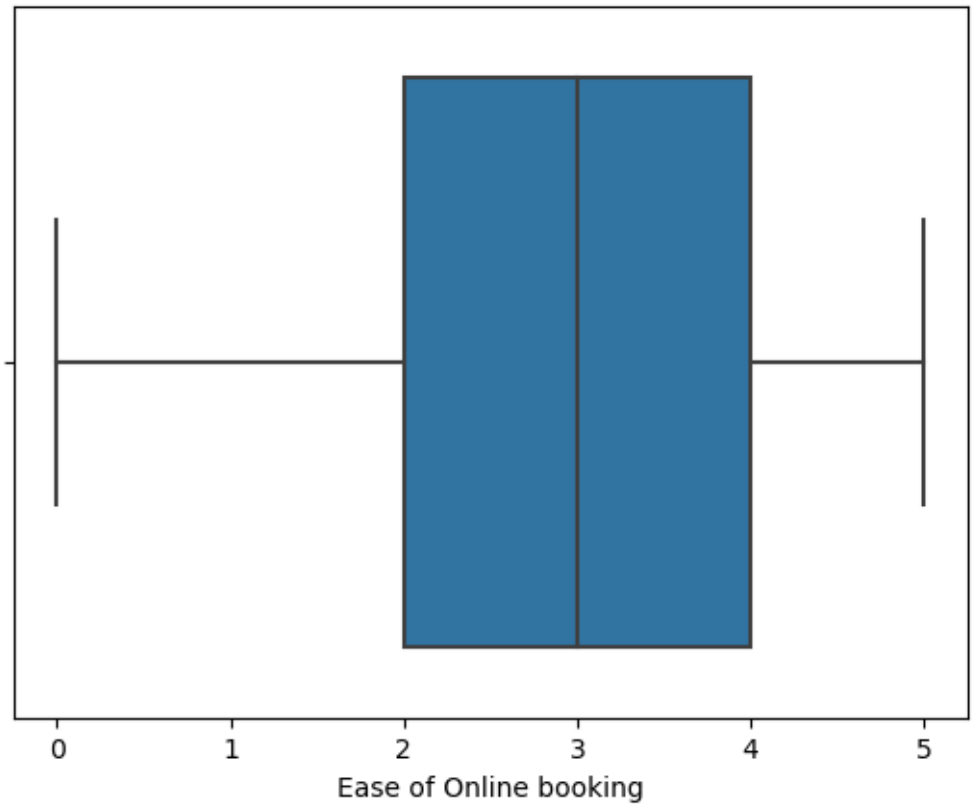

0
20000
40000
60000
80000
100000
120000
id

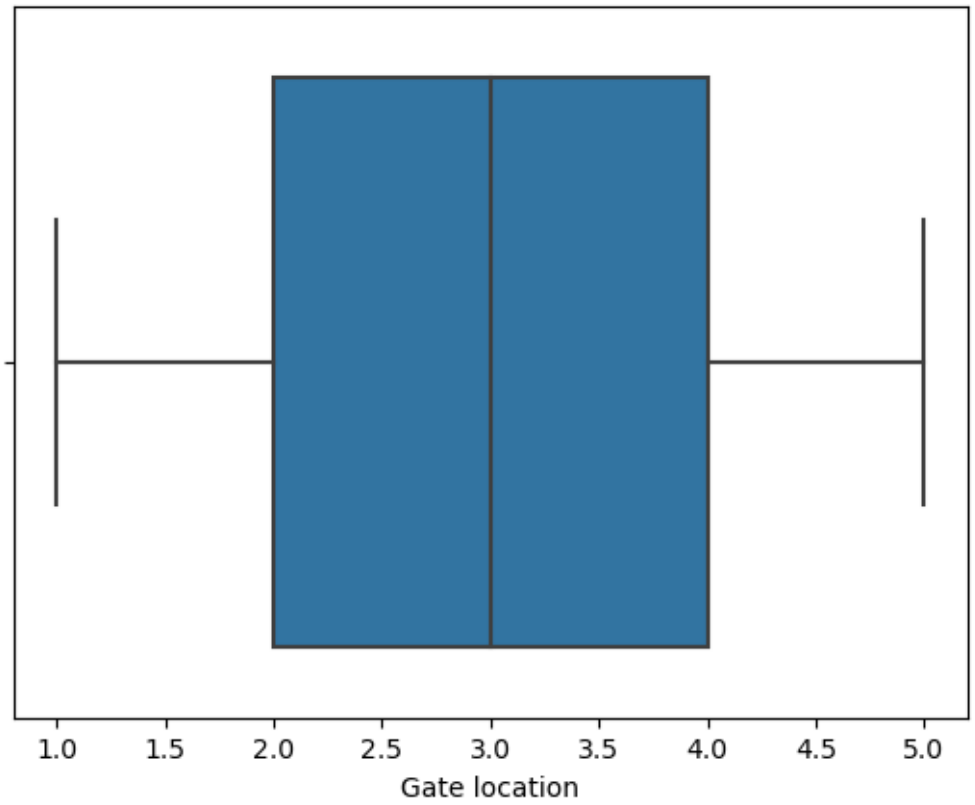

10
20
30
40
50
60
70
80
Age

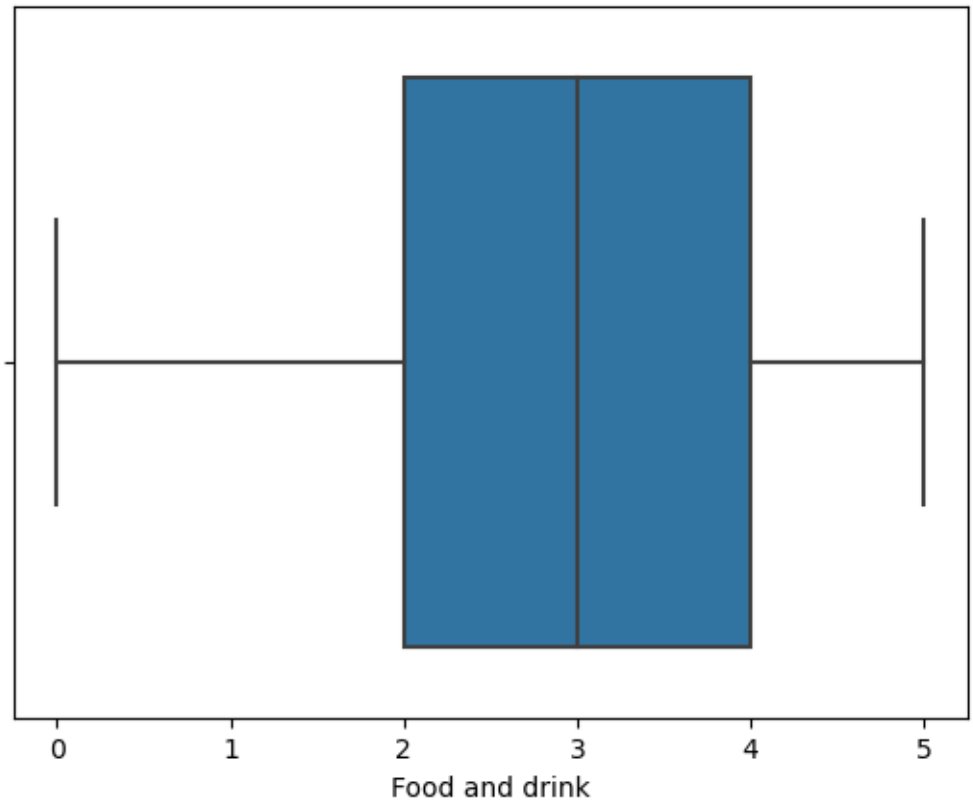

0
1000
2000
3000
4000
5000
Flight Distance

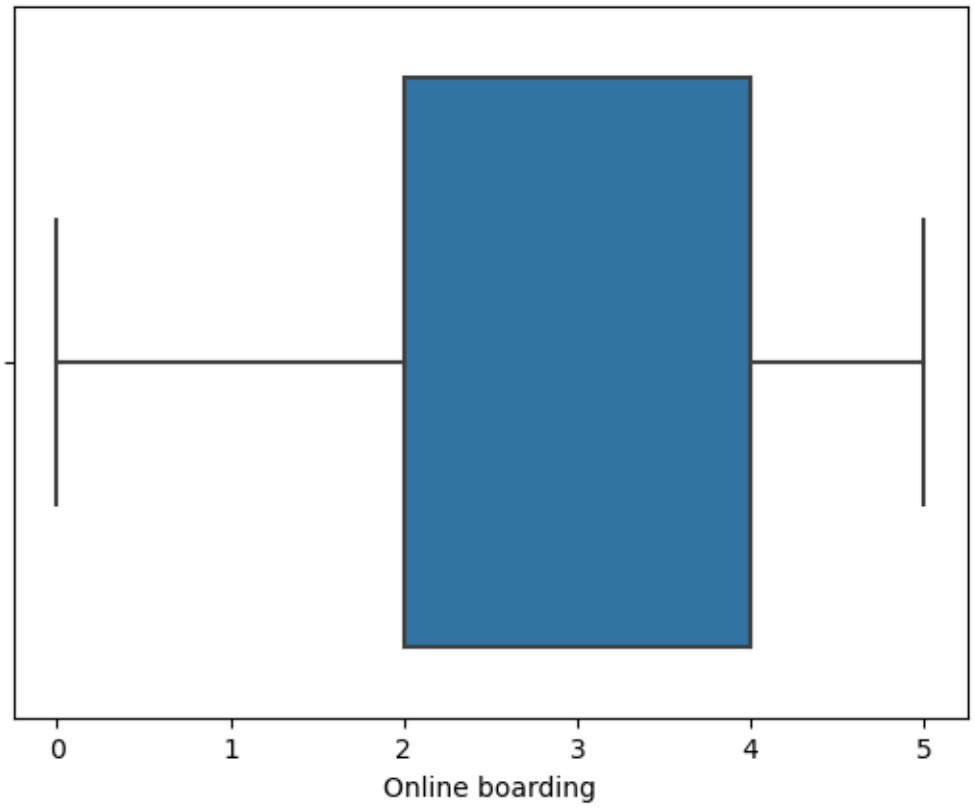

0
1
2
3
4
5
Inflight wifi service

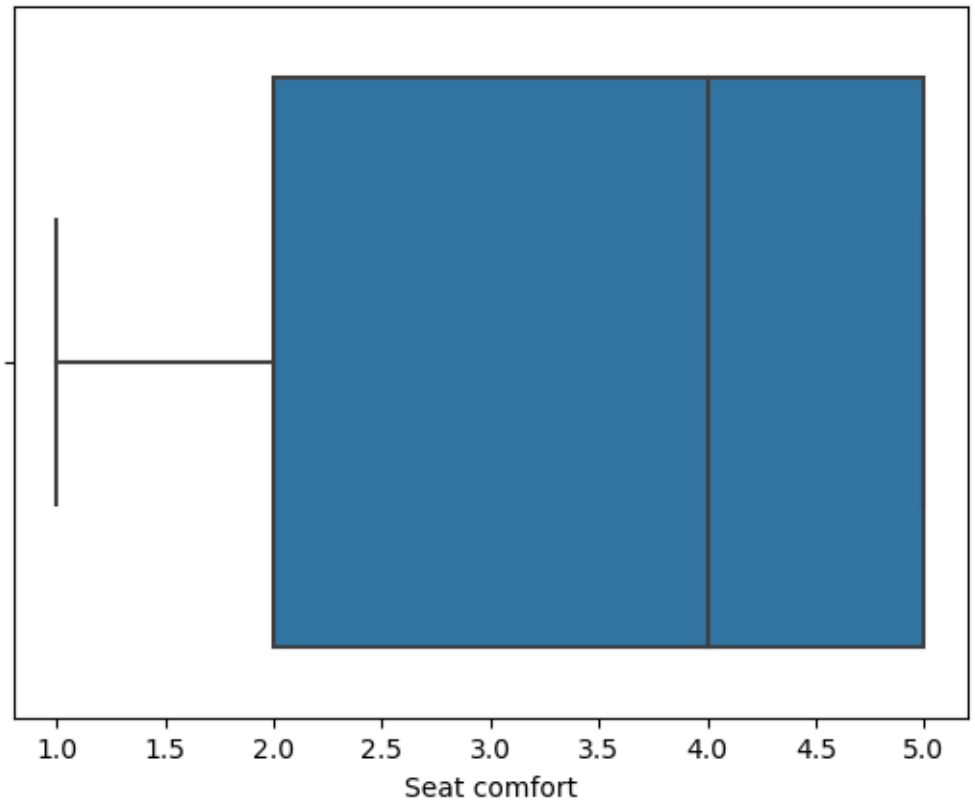

0
1
2
3
4
5
Departure_Arrival time convenient

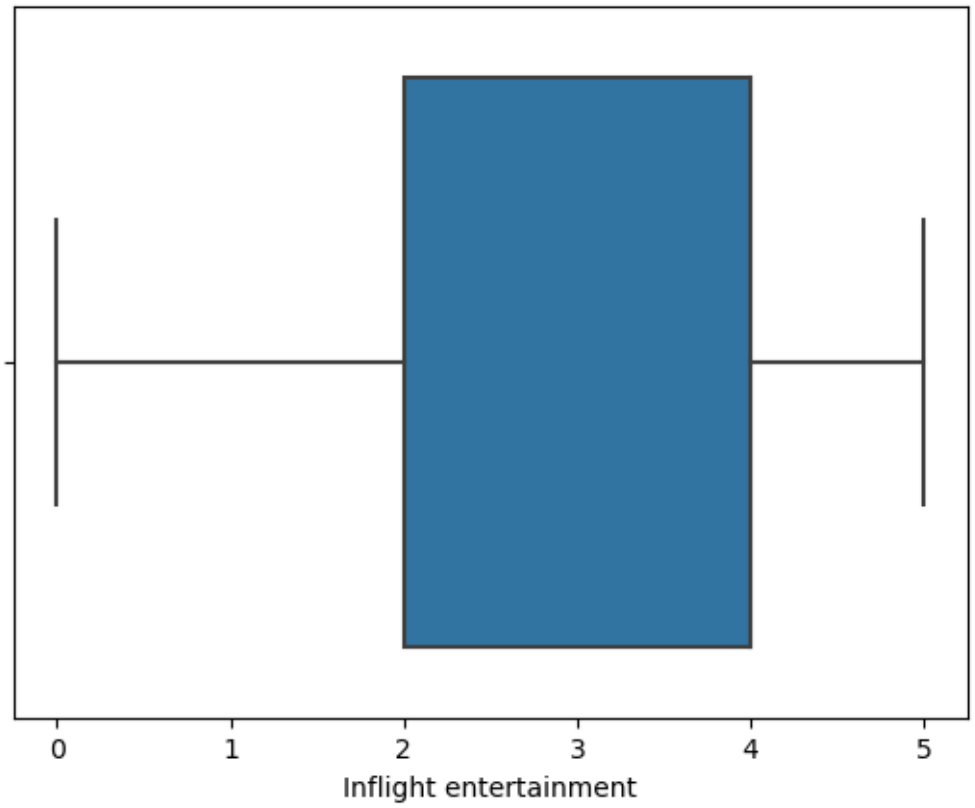

0
1
2
3
4
5
Ease of Online booking

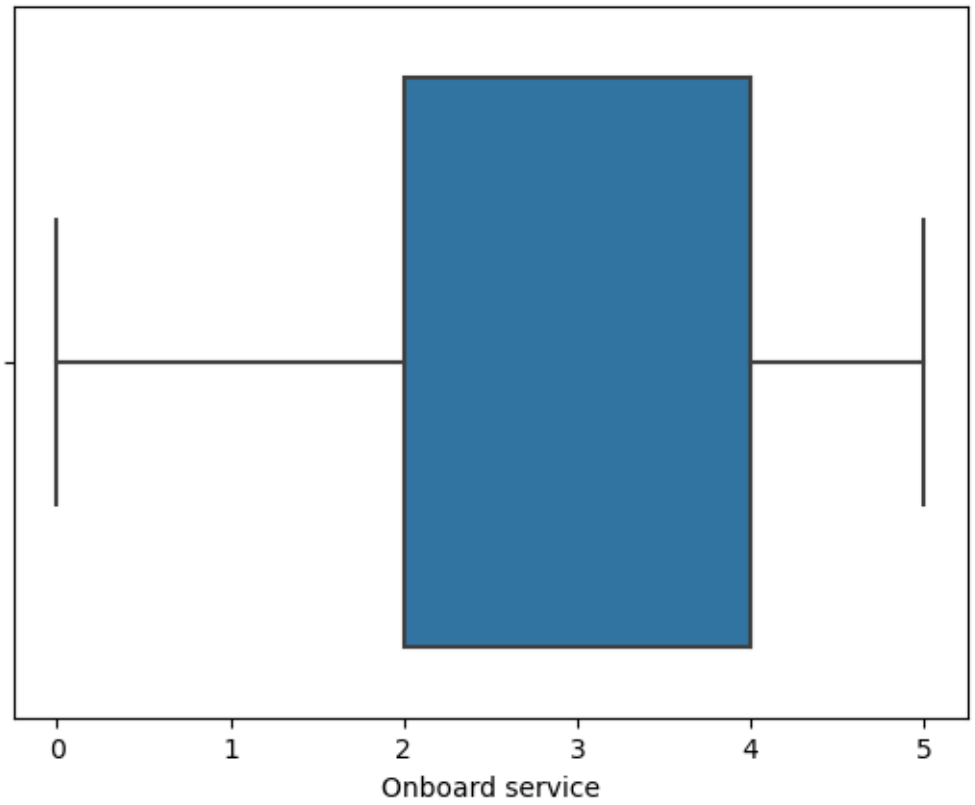

1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Gate location

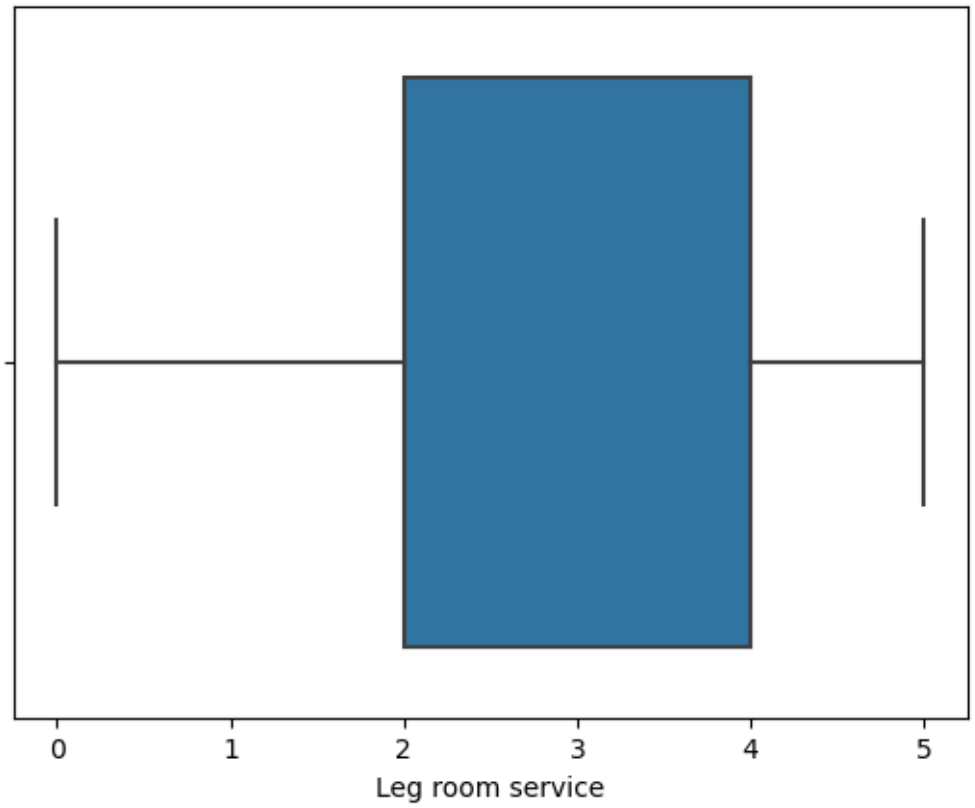

0
1
2
3
4
5
Food and drink

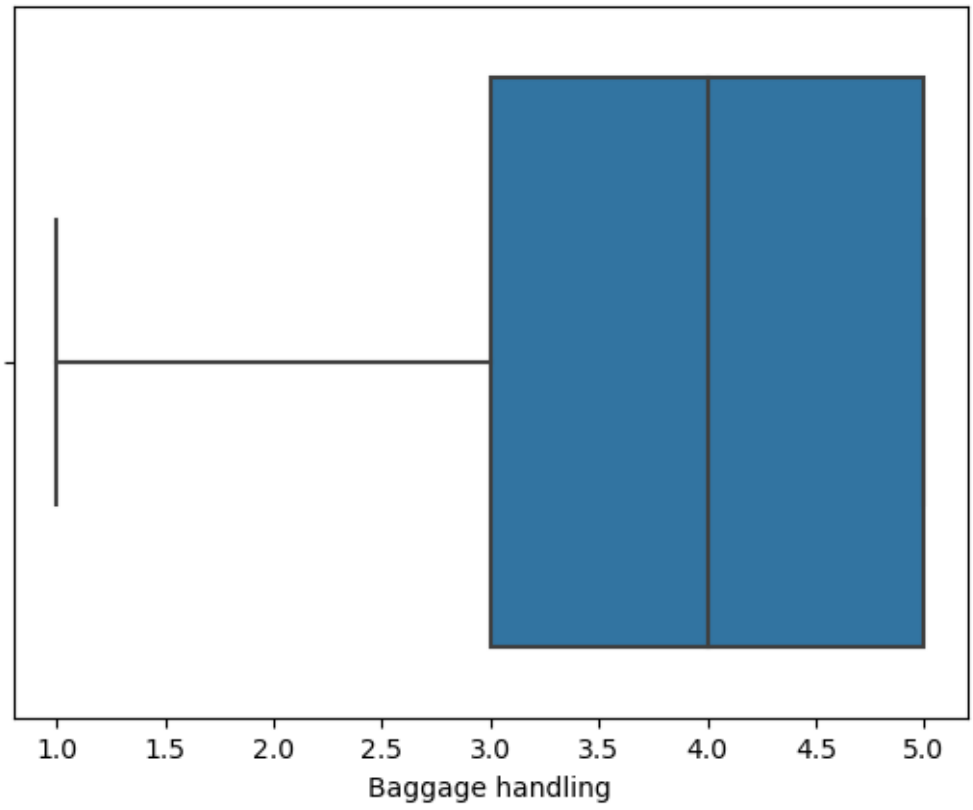

0
1
2
3
4
5
Online boarding

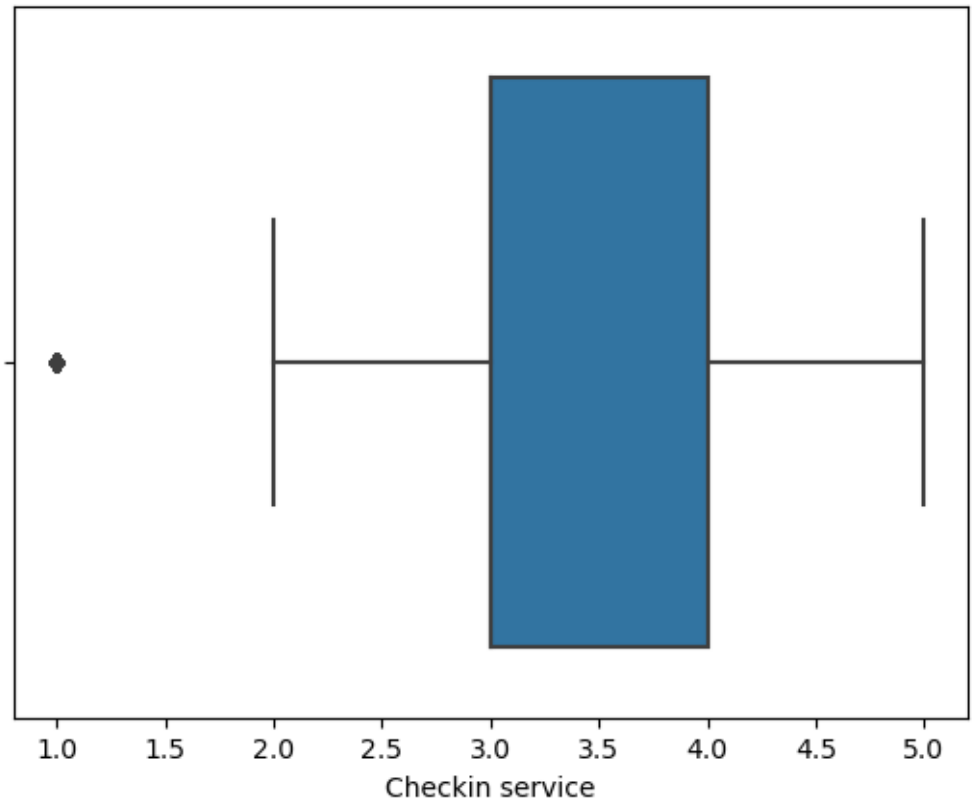

1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Seat comfort

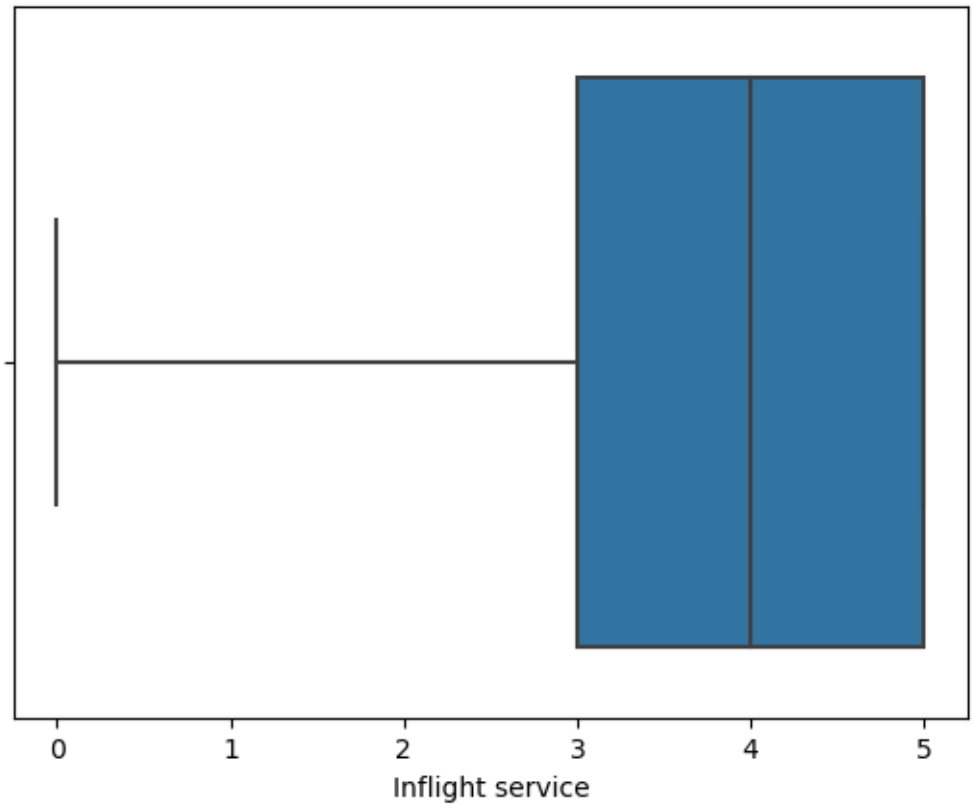

0
1
2
3
4
5
Inflight entertainment

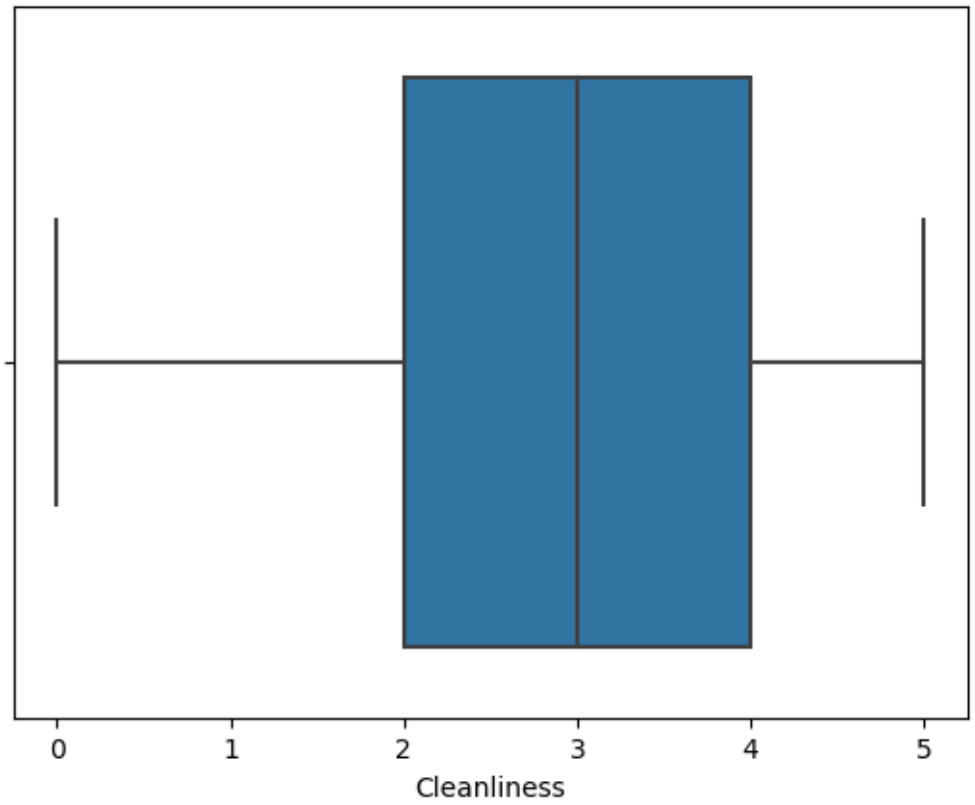

0
1
2
3
4
5
Onboard service

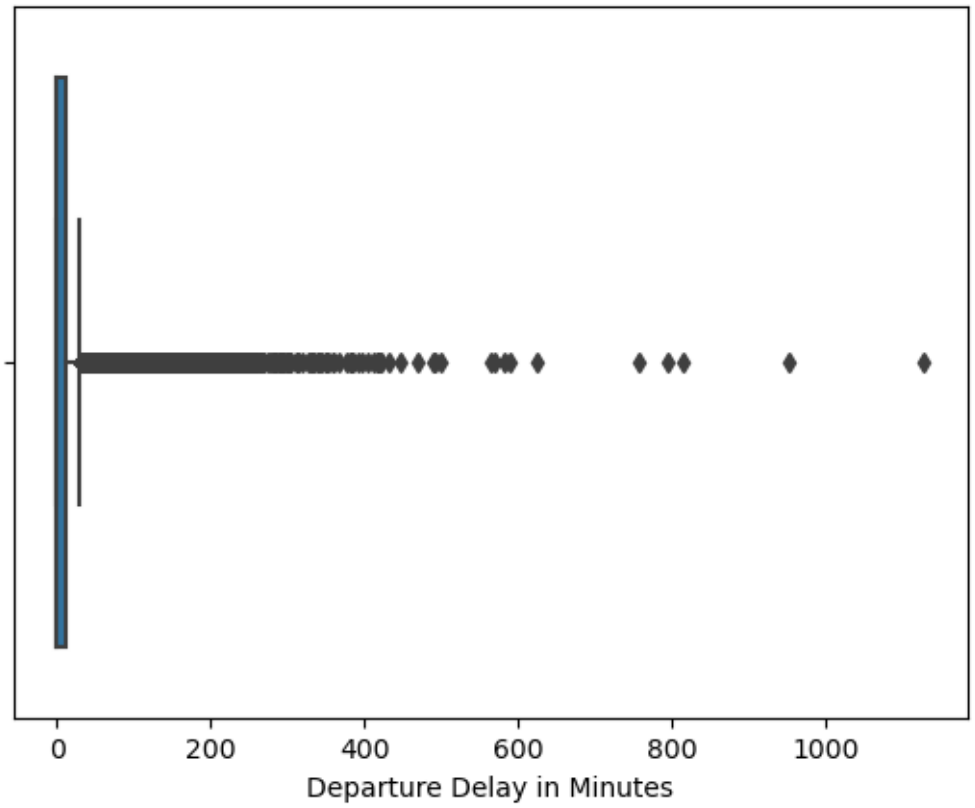

0
1
2
3
4
5
Leg room service

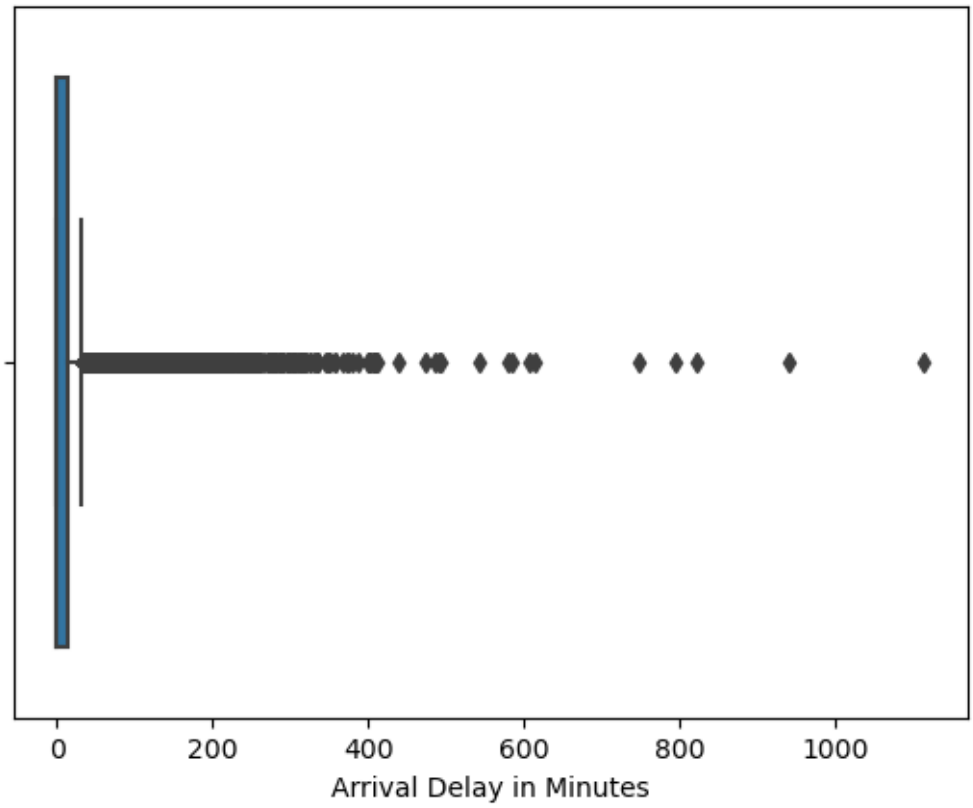

1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Baggage handling


1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Checkin service


0
1
2
3
4
5
Inflight service

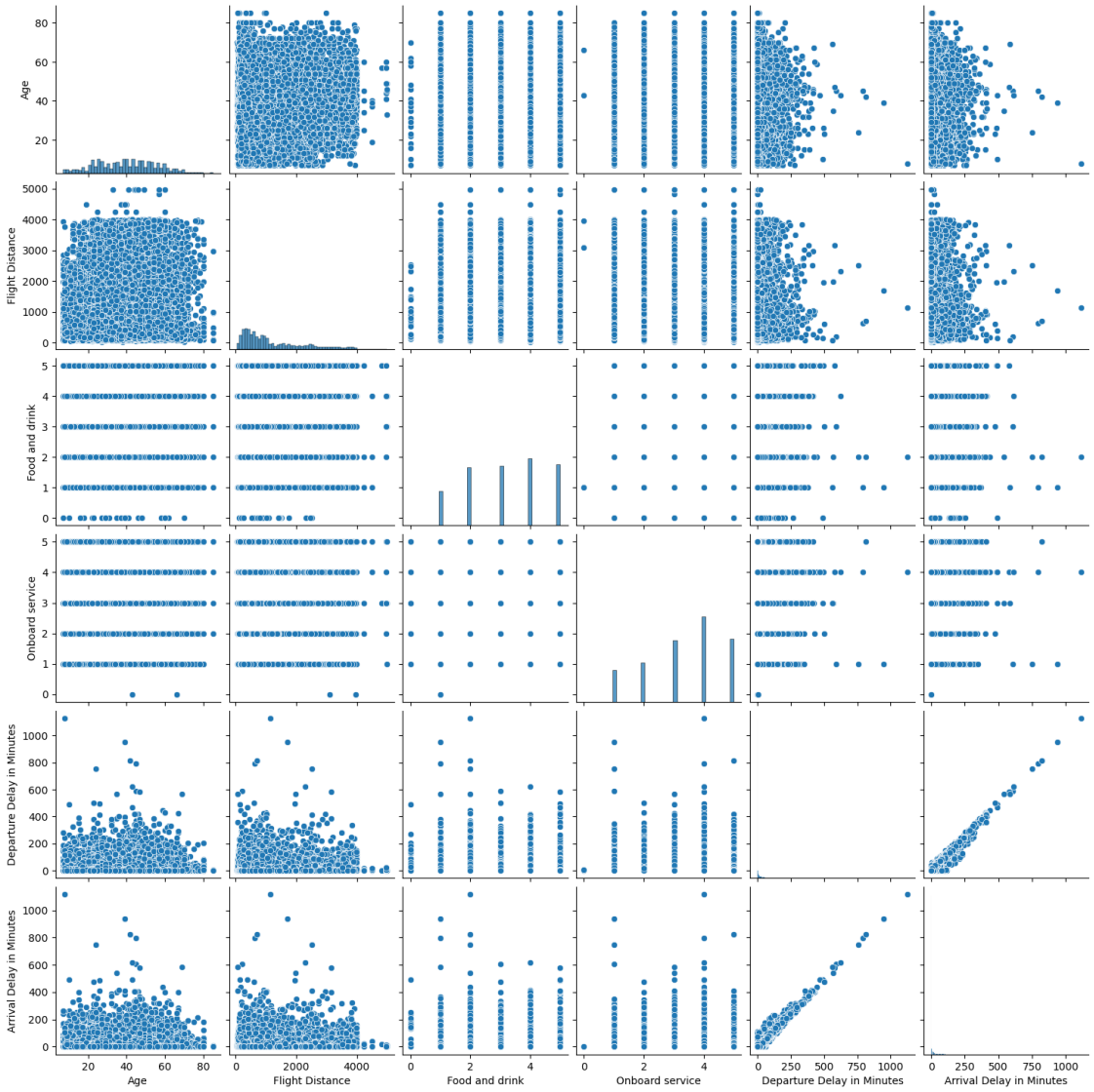

0
1
2
3
4
5
Cleanliness


0
200
400
600
800
1000
Departure Delay in Minutes



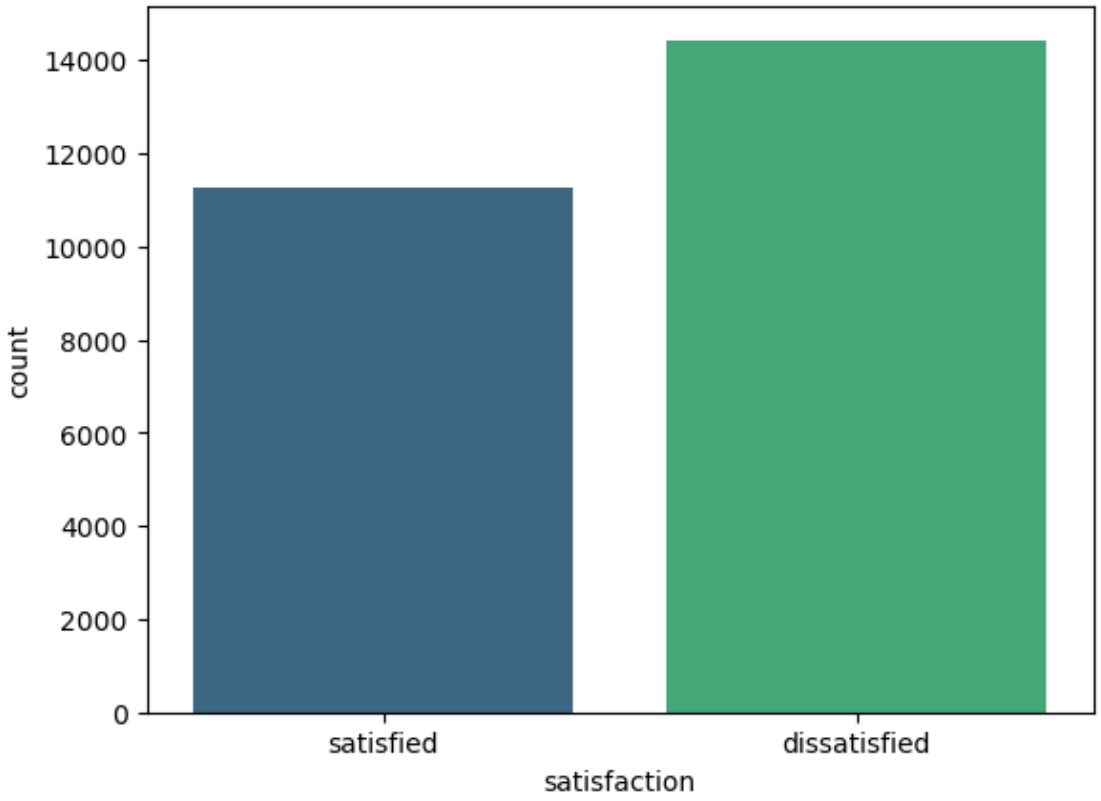
[31]:
```
pair_columns=['Age','Flight Distance','Food and drink','Onboard␣
 ↪service','Departure Delay in Minutes','Arrival Delay in Minutes']
sns.pairplot(df[pair_columns])
```

[31]: <seaborn.axisgrid.PairGrid at 0x160ce9a8b50>



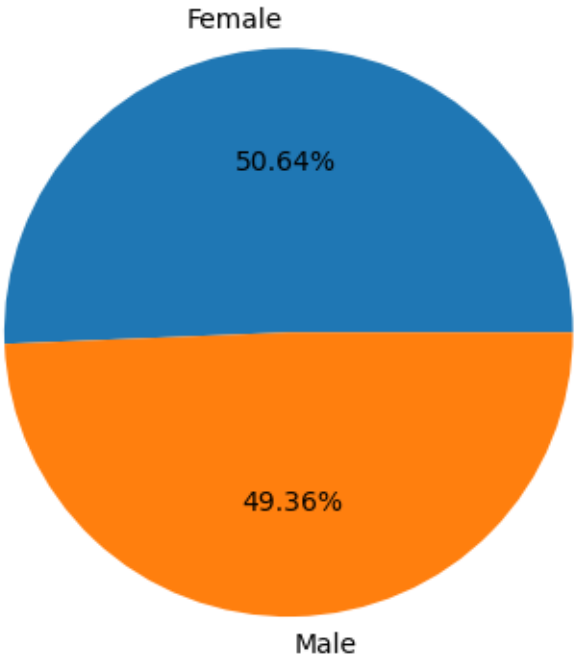
[32]:
```
sns.countplot(x='satisfaction', data=df, palette='viridis')
```

[32]: <Axes: xlabel='satisfaction', ylabel='count'>



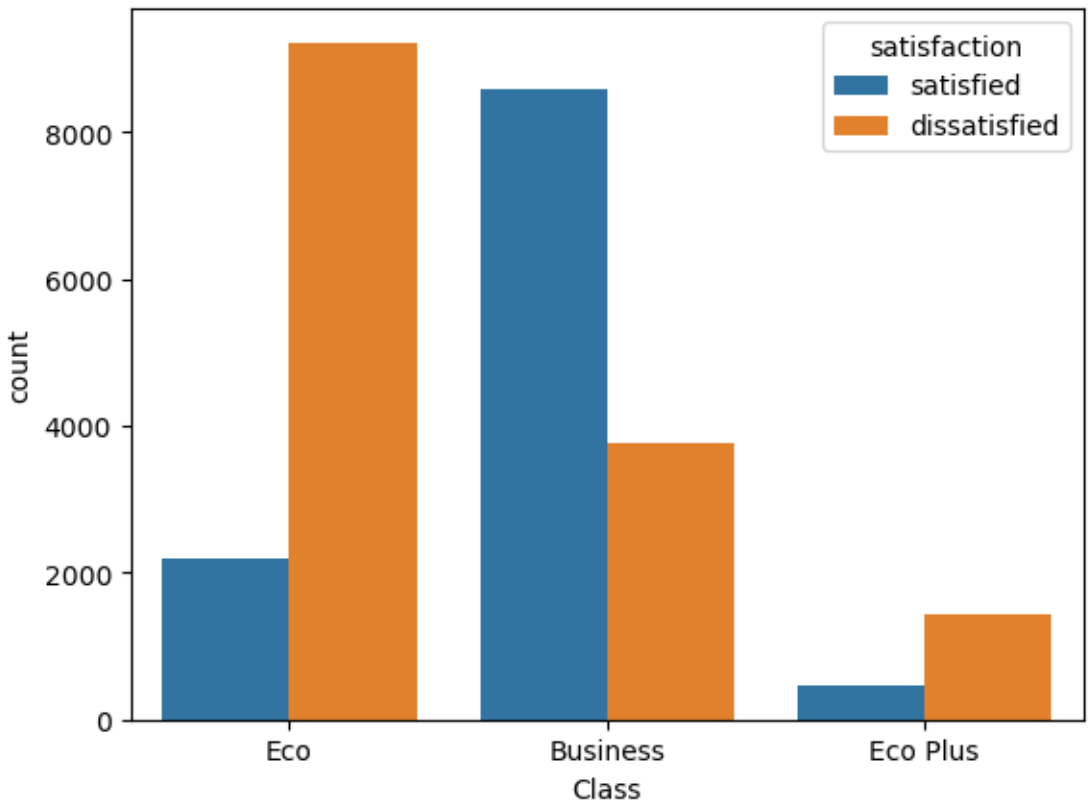
[33]:
```
a=df.groupby('Gender')['Gender'].count()
plt.pie(a,labels=a.index,autopct='%.2f%%')
```

[33]:
```
([<matplotlib.patches.Wedge at 0x160d71d6f50>,
  <matplotlib.patches.Wedge at 0x160d7c1c4d0>],
 [Text(-0.022260264214290826, 1.0997747408615595, 'Female'),
  Text(0.022260264214290445, -1.0997747408615595, 'Male')],
 [Text(-0.012141962298704085, 0.5998771313790323, '50.64%'),
  Text(0.012141962298703877, -0.5998771313790323, '49.36%')])
```



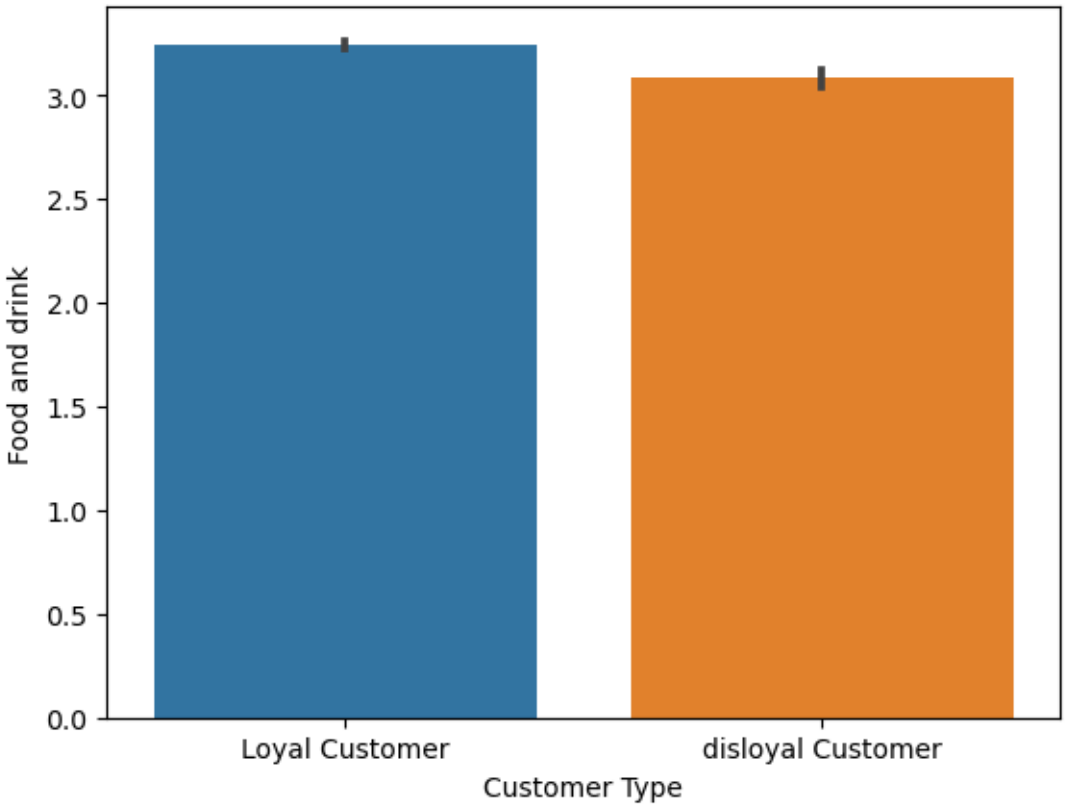
[34]:
```
sns.countplot(x='Class',data=df,hue='satisfaction')
```

[34]:
```
<Axes: xlabel='Class', ylabel='count'>
```



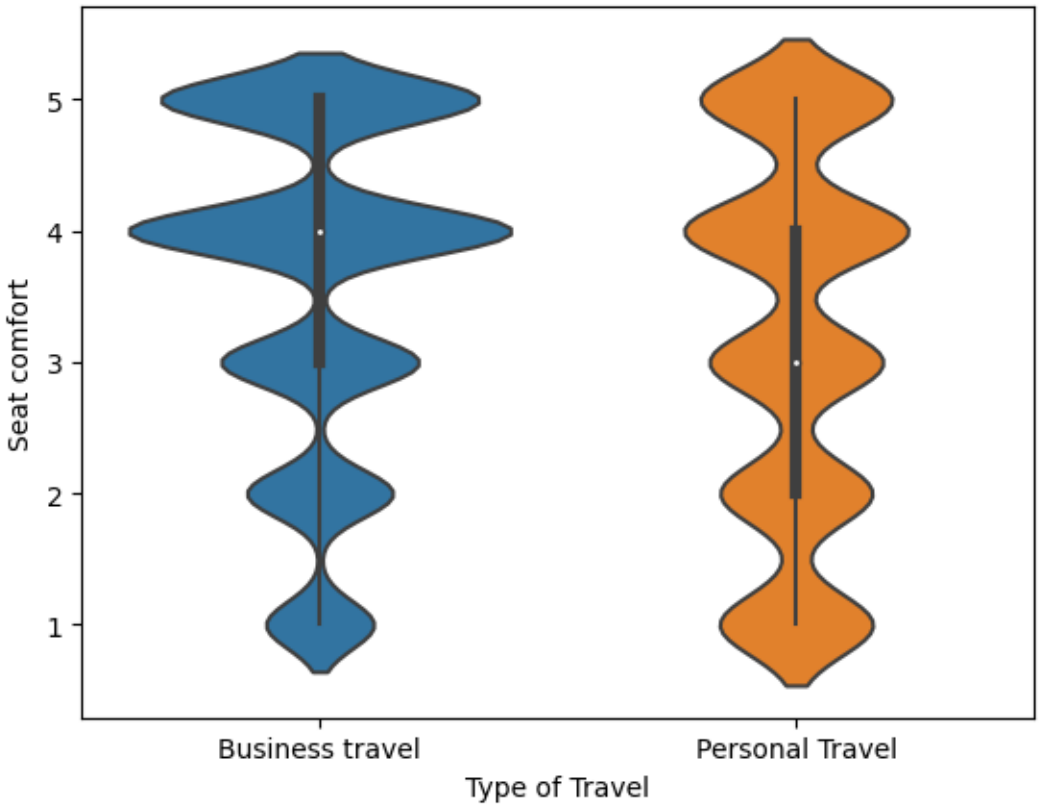
[35]:
```
sns.barplot(x='Customer Type',y='Food and drink',data=df)
```

[35]:
```
<Axes: xlabel='Customer Type', ylabel='Food and drink'>
```



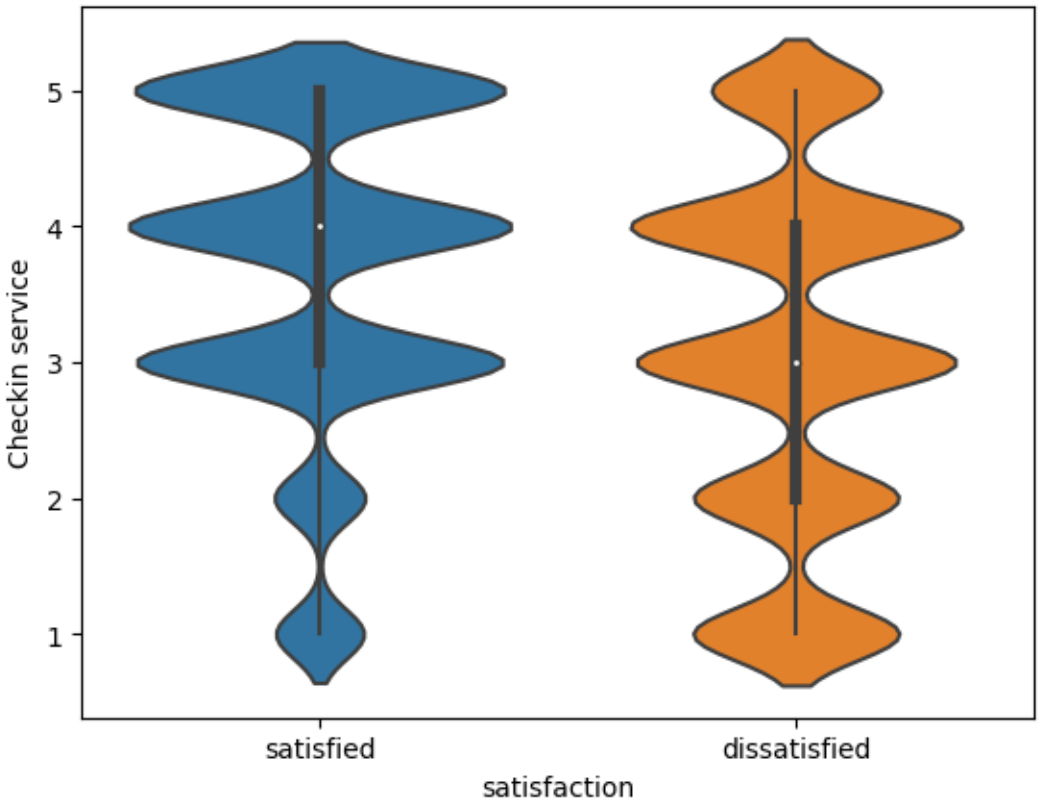
[36]:
```
sns.violinplot(x='Type of Travel',y='Seat comfort',data=df)
```

[36]: <Axes: xlabel='Type of Travel', ylabel='Seat comfort'>



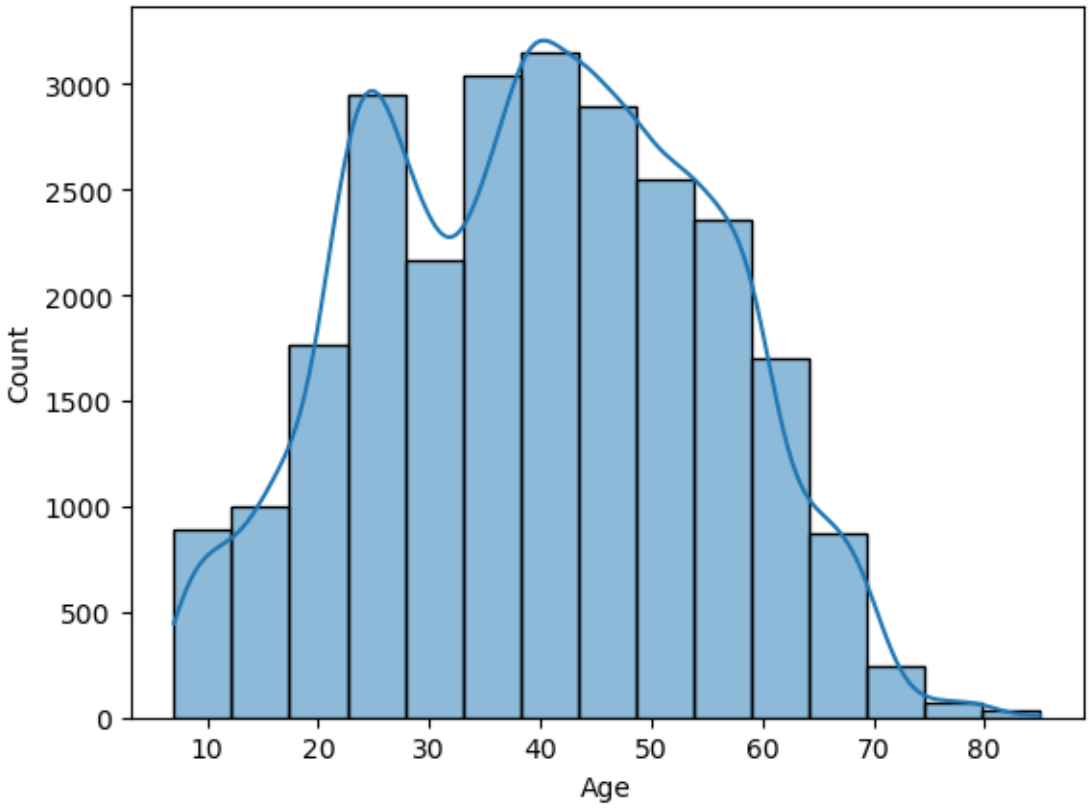
[37]:
```
sns.violinplot(x='satisfaction',y='Checkin service',data=df)
```

[37]:
```
<Axes: xlabel='satisfaction', ylabel='Checkin service'>
```



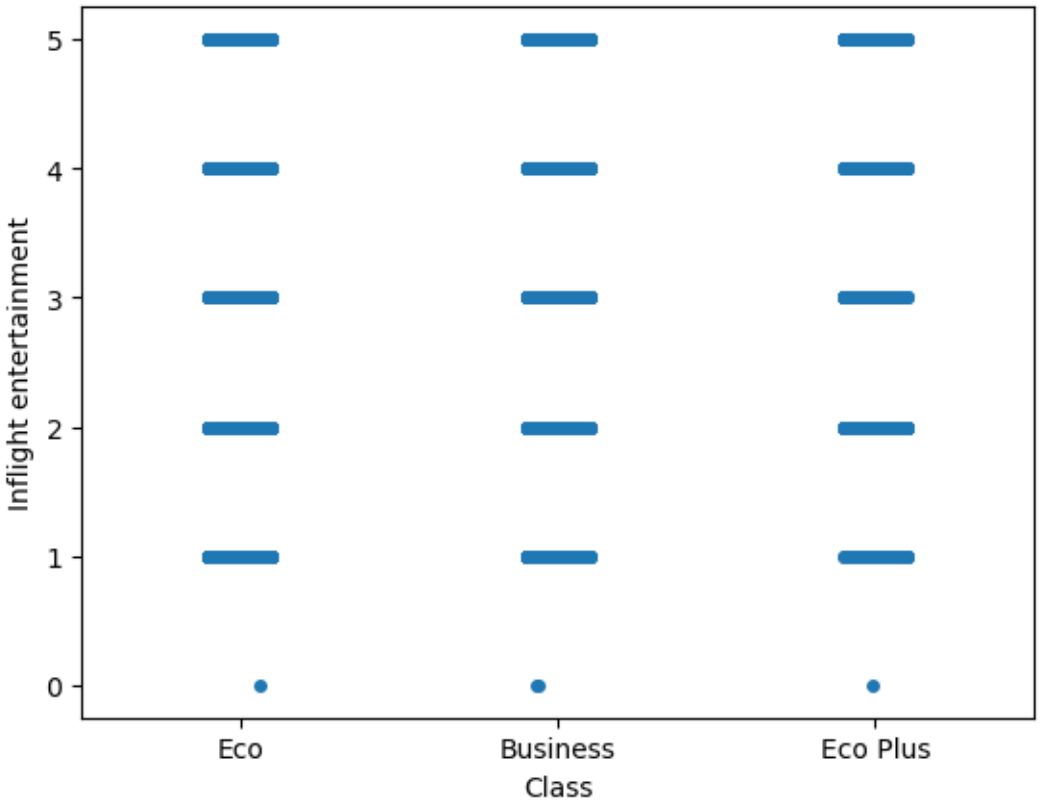
[38]:
```
sns.histplot(x='Age',data=df,bins=15,kde=True)
```

[38]:
```
<Axes: xlabel='Age', ylabel='Count'>
```



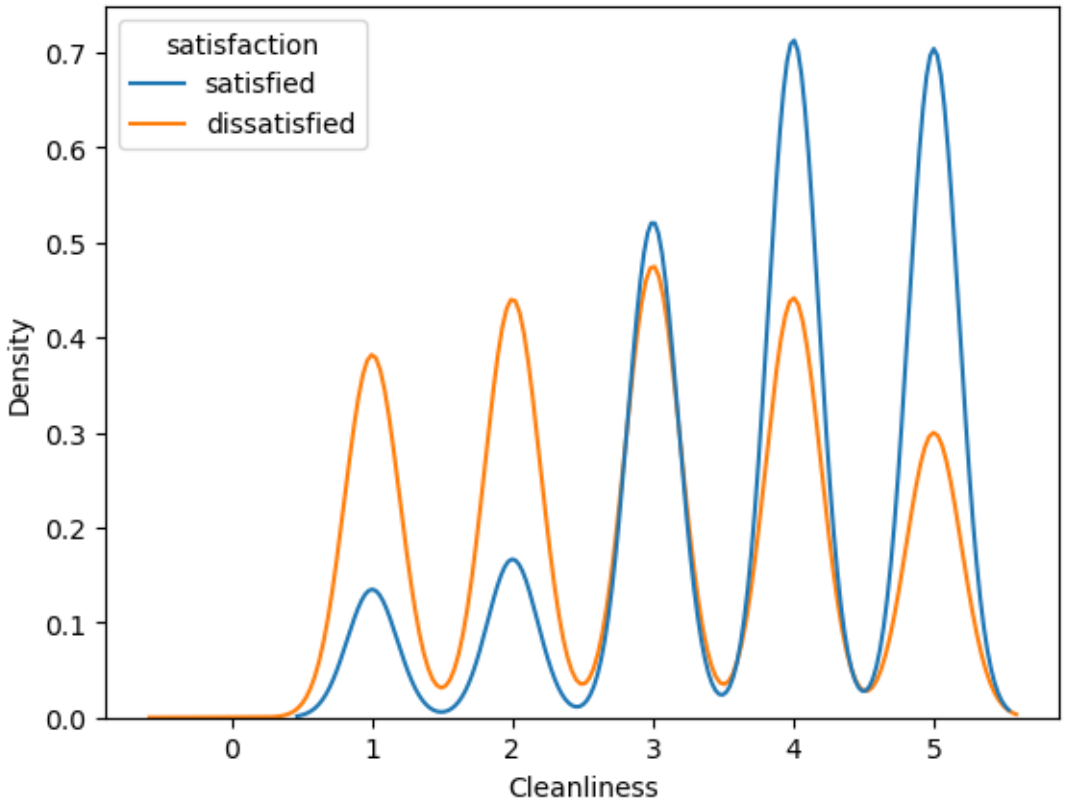
[39]:
```
sns.stripplot(x='Class',y='Inflight entertainment',data=df)
```

[39]:
```
<Axes: xlabel='Class', ylabel='Inflight entertainment'>
```



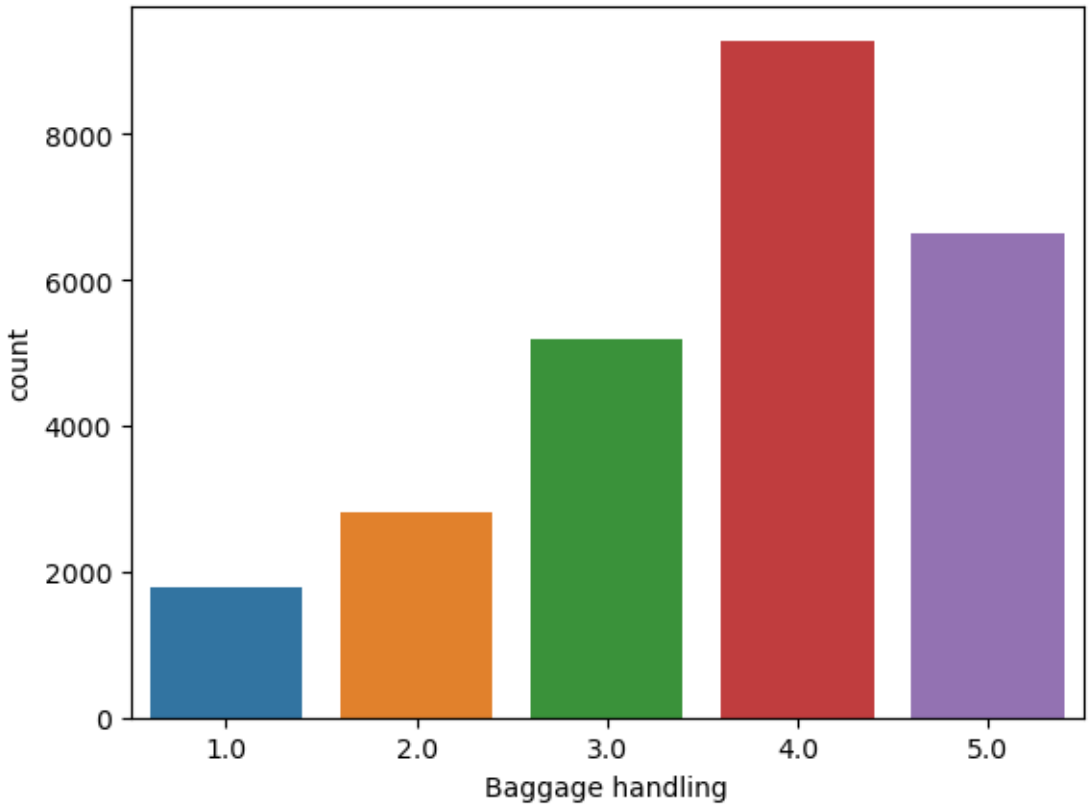
[40]:
```
sns.kdeplot(data=df, x='Cleanliness', hue='satisfaction',␣
↪common_norm=False,warn_singular=False)
```

[40]: <Axes: xlabel='Cleanliness', ylabel='Density'>



[41]:
```
sns.countplot(x='Baggage handling',data=df)
```

[41]:
```
<Axes: xlabel='Baggage handling', ylabel='count'>
```



[42]:
```
from scipy import stats
z_scores=stats.zscore(df['Flight Distance'])
z_score_outliers=(z_scores<-3)|(z_scores>3)
z_score_outlier_rows=df[z_score_outliers]
print("outliers detected by Z-score:",z_score_outlier_rows)
```

```
outliers detected by Z-score:        sr_no     id  Gender      Customer Type
Age   Type of Travel  \
273       273  53258  Female     Loyal Customer   57  Business travel
3645     3645  59263    Male     Loyal Customer   44  Business travel
3795     3795  72042  Female     Loyal Customer   45  Personal Travel
3953     3953  59265    Male     Loyal Customer   44  Business travel
7851     7851  31874  Female     Loyal Customer   33  Business travel
11779   11779  31816  Female     Loyal Customer   45  Personal Travel
13464   13464  58700    Male     Loyal Customer   60  Business travel
13666   13666  58706  Female  disloyal Customer   25  Business travel
14261   14261  31872    Male     Loyal Customer   46  Business travel
14591   14591  56482    Male     Loyal Customer   41  Personal Travel
14615   14615  31815  Female     Loyal Customer   49  Business travel
15940   15940  87733    Male     Loyal Customer   19  Business travel
16795   16795  56477    Male     Loyal Customer   57  Business travel
```

```
21026  21026  59272    Male     Loyal Customer   34  Business travel
22817  22817  58702    Male     Loyal Customer   40  Business travel
23360  23360  56476    Male     Loyal Customer   60  Business travel
24448  24448  84495    Male  disloyal Customer   40  Business travel
24452  24452  84501  Female     Loyal Customer   37  Business travel
24940  24940  84497  Female     Loyal Customer   39  Business travel
25704  25704  59264    Male     Loyal Customer   49  Business travel

         Class  Flight Distance  Inflight wifi service  \
273    Business           4817.0                    4.0
3645   Business           4963.0                    5.0
3795        Eco           4243.0                    1.0
3953   Business           4963.0                    5.0
7851   Business           4983.0                    1.0
11779       Eco           4983.0                    3.0
13464  Business           4243.0                    3.0
13666  Business           4243.0                    1.0
14261  Business           4983.0                    1.0
14591       Eco           4963.0                    2.0
14615  Business           4983.0                    4.0
15940  Business           4502.0                    2.0
16795  Business           4963.0                    4.0
21026       Eco           4243.0                    5.0
22817  Business           4243.0                    2.0
23360  Business           4963.0                    2.0
24448  Business           4502.0                    4.0
24452  Business           4502.0                    2.0
24940       Eco           4502.0                    1.0
25704  Business           4963.0                    1.0

       Departure_Arrival time convenient  …  Inflight entertainment  \
273                                  4.0  …                     3.0
3645                                 5.0  …                     3.0
3795                                 5.0  …                     1.0
3953                                 5.0  …                     3.0
7851                                 1.0  …                     4.0
11779                                1.0  …                     2.0
13464                                3.0  …                     3.0
13666                                4.0  …                     3.0
14261                                1.0  …                     1.0
14591                                4.0  …                     4.0
14615                                1.0  …                     2.0
15940                                1.0  …                     3.0
16795                                4.0  …                     5.0
21026                                5.0  …                     4.0
22817                                5.0  …                     4.0
23360                                5.0  …                     5.0
24448                                4.0  …                     4.0
```

```
24452                          2.0  …                  4.0
24940                          2.0  …                  2.0
25704                          1.0  …                  4.0

       Onboard service  Leg room service  Baggage handling  Checkin service  \
273                3.0               4.0               4.0              5.0
3645               3.0               4.0               4.0              4.0
3795               2.0               3.0               3.0              3.0
3953               2.0               3.0               4.0              4.0
7851               1.0               4.0               5.0              4.0
11779              5.0               2.0               3.0              1.0
13464              5.0               3.0               5.0              5.0
13666              2.0               3.0               5.0              3.0
14261              3.0               5.0               5.0              4.0
14591              5.0               4.0               3.0              3.0
14615              5.0               2.0               3.0              4.0
15940              5.0               1.0               2.0              2.0
16795              3.0               4.0               4.0              3.0
21026              5.0               2.0               5.0              5.0
22817              3.0               1.0               3.0              2.0
23360              3.0               4.0               4.0              3.0
24448              4.0               4.0               4.0              4.0
24452              5.0               5.0               4.0              4.0
24940              4.0               4.0               4.0              1.0
25704              4.0               3.0               4.0              4.0

       Inflight service  Cleanliness  Departure Delay in Minutes  \
273                 4.0          5.0                         0.0
3645                5.0          4.0                         0.0
3795                2.0          3.0                         6.0
3953                4.0          4.0                         0.0
7851                3.0          4.0                         0.0
11779               4.0          1.0                         2.0
13464               4.0          5.0                        18.0
13666               4.0          3.0                         0.0
14261               3.0          4.0                         0.0
14591               2.0          3.0                         0.0
14615               4.0          4.0                         0.0
15940               2.0          2.0                        12.0
16795               4.0          3.0                         4.0
21026               5.0          5.0                         4.0
22817               1.0          2.0                         0.0
23360               3.0          3.0                         5.0
24448               5.0          4.0                         0.0
24452               3.0          4.0                         6.0
24940               4.0          1.0                        16.0
25704               4.0          4.0                        23.0
```

```
       Arrival Delay in Minutes  satisfaction
273                        19.0     satisfied
3645                        0.0     satisfied
3795                        0.0  dissatisfied
3953                        0.0     satisfied
7851                        0.0     satisfied
11779                       0.0  dissatisfied
13464                      23.0     satisfied
13666                       1.0  dissatisfied
14261                       0.0     satisfied
14591                       4.0  dissatisfied
14615                      14.0  dissatisfied
15940                      42.0  dissatisfied
16795                      13.0     satisfied
21026                       1.0     satisfied
22817                      19.0  dissatisfied
23360                      13.0     satisfied
24448                       0.0  dissatisfied
24452                       0.0     satisfied
24940                       0.0  dissatisfied
25704                       0.0     satisfied

[20 rows x 25 columns]
```

[43]:
```
x=(z_scores>-3)&(z_scores<3)
df_new=df[x]
```

[44]:
```
z_scores=stats.zscore(df['Departure Delay in Minutes'])
z_score_outliers=(z_scores<-3)|(z_scores>3)
z_score_outlier_rows=df[z_score_outliers]
print("outliers detected by Z-score:",z_score_outlier_rows)
```

```
outliers detected by Z-score:        sr_no      id  Gender      Customer Type
Age   Type of Travel  \
145      145    7237  Female     Loyal Customer   53  Business travel
279      279   24628  Female     Loyal Customer   14  Personal Travel
281      281   62793    Male     Loyal Customer   11  Personal Travel
337      337  113849    Male     Loyal Customer   50  Business travel
345      345   35387  Female     Loyal Customer   53  Business travel
…        …      …      …                 …  …               …
25686  25686   20159    Male     Loyal Customer   61  Business travel
25697  25697   48833  Female     Loyal Customer   70  Personal Travel
25757  25757   38373  Female     Loyal Customer   58  Personal Travel
25771  25771   15343    Male  disloyal Customer   44  Business travel
25942  25942   91254    Male     Loyal Customer   44  Personal Travel

          Class  Flight Distance  Inflight wifi service  \
145    Business           3769.0                    3.0
```

```
279         Eco            666.0                    3.0
281         Eco           2338.0                    2.0
337    Business           1363.0                    2.0
345    Business            635.0                    2.0
…         …                …                        …
25686  Business            403.0                    2.0
25697  Business            266.0                    1.0
25757       Eco           1133.0                    1.0
25771  Business            331.0                    2.0
25942       Eco            404.0                    3.0

       Departure_Arrival time convenient  …  Inflight entertainment  \
145                                  2.0  …                     2.0
279                                  3.0  …                     3.0
281                                  4.0  …                     3.0
337                                  2.0  …                     3.0
345                                  2.0  …                     1.0
…                                    …  …                     …
25686                                4.0  …                     1.0
25697                                4.0  …                     3.0
25757                                3.0  …                     3.0
25771                                2.0  …                     5.0
25942                                1.0  …                     2.0

       Onboard service  Leg room service  Baggage handling  Checkin service  \
145                4.0               3.0               4.0              1.0
279                4.0               4.0               5.0              1.0
281                2.0               4.0               5.0              1.0
337                5.0               4.0               4.0              2.0
345                1.0               2.0               3.0              5.0
…                  …                 …                 …                …
25686              4.0               3.0               4.0              2.0
25697              2.0               4.0               3.0              1.0
25757              1.0               3.0               3.0              1.0
25771              2.0               4.0               5.0              5.0
25942              4.0               2.0               3.0              2.0

       Inflight service  Cleanliness  Departure Delay in Minutes  \
145                 2.0          1.0                       180.0
279                 4.0          1.0                       243.0
281                 4.0          1.0                       152.0
337                 5.0          2.0                       238.0
345                 3.0          5.0                       150.0
…                   …            …                         …
25686               3.0          2.0                       136.0
25697               1.0          1.0                       153.0
25757               1.0          1.0                       151.0
25771               5.0          5.0                       295.0
```

```
25942               4.0          2.0                       131.0

       Arrival Delay in Minutes  satisfaction
145                       178.0  dissatisfied
279                       251.0  dissatisfied
281                       151.0  dissatisfied
337                       232.0     satisfied
345                       156.0     satisfied
…                             …             …
25686                     155.0  dissatisfied
25697                     143.0  dissatisfied
25757                     162.0  dissatisfied
25771                     288.0  dissatisfied
25942                     138.0  dissatisfied

[502 rows x 25 columns]
```

[45]:
```
x=(z_scores>-3)&(z_scores<3)
df_new=df[x]
```

[46]:
```
z_scores=stats.zscore(df['Arrival Delay in Minutes'])
z_score_outliers=(z_scores<-3)|(z_scores>3)
z_score_outlier_rows=df[z_score_outliers]
print("outliers detected by Z-score:",z_score_outlier_rows)
x=(z_scores>-3)&(z_scores<3)
df_new=df[x]
```

```
outliers detected by Z-score:        sr_no      id  Gender      Customer Type
Age   Type of Travel  \
42         42   16172    Male     Loyal Customer   22  Business travel
145       145    7237  Female     Loyal Customer   53  Business travel
279       279   24628  Female     Loyal Customer   14  Personal Travel
281       281   62793    Male     Loyal Customer   11  Personal Travel
337       337  113849    Male     Loyal Customer   50  Business travel
…         …     …     …                  …  …               …
25697  25697   48833  Female     Loyal Customer   70  Personal Travel
25757  25757   38373  Female     Loyal Customer   58  Personal Travel
25771  25771   15343    Male  disloyal Customer   44  Business travel
25904  25904   89801  Female     Loyal Customer   49  Business travel
25942  25942   91254    Male     Loyal Customer   44  Personal Travel

          Class  Flight Distance  Inflight wifi service  \
42     Business            277.0                    3.0
145    Business           3769.0                    3.0
279         Eco            666.0                    3.0
281         Eco           2338.0                    2.0
337    Business           1363.0                    2.0
…         …              …                      …
```

```
25697  Business            266.0                   1.0
25757       Eco           1133.0                   1.0
25771  Business            331.0                   2.0
25904  Business           3542.0                   3.0
25942       Eco            404.0                   3.0

       Departure_Arrival time convenient  …  Inflight entertainment  \
42                                   3.0  …                     1.0
145                                  2.0  …                     2.0
279                                  3.0  …                     3.0
281                                  4.0  …                     3.0
337                                  2.0  …                     3.0
…                                    …  …                     …
25697                                4.0  …                     3.0
25757                                3.0  …                     3.0
25771                                2.0  …                     5.0
25904                                3.0  …                     3.0
25942                                1.0  …                     2.0

       Onboard service  Leg room service  Baggage handling  Checkin service  \
42                 5.0                4.0               3.0              3.0
145                4.0                3.0               4.0              1.0
279                4.0                4.0               5.0              1.0
281                2.0                4.0               5.0              1.0
337                5.0                4.0               4.0              2.0
…                  …                  …                 …                …
25697              2.0                4.0               3.0              1.0
25757              1.0                3.0               3.0              1.0
25771              2.0                4.0               5.0              5.0
25904              4.0                5.0               5.0              5.0
25942              4.0                2.0               3.0              2.0

       Inflight service  Cleanliness  Departure Delay in Minutes  \
42                  2.0          3.0                       116.0
145                 2.0          1.0                       180.0
279                 4.0          1.0                       243.0
281                 4.0          1.0                       152.0
337                 5.0          2.0                       238.0
…                   …            …                         …
25697               1.0          1.0                       153.0
25757               1.0          1.0                       151.0
25771               5.0          5.0                       295.0
25904               5.0          5.0                        90.0
25942               4.0          2.0                       131.0

       Arrival Delay in Minutes  satisfaction
42                        177.0  dissatisfied
145                       178.0  dissatisfied
```

```
279                       251.0  dissatisfied
281                       151.0  dissatisfied
337                       232.0     satisfied
…                           …           …
25697                     143.0  dissatisfied
25757                     162.0  dissatisfied
25771                     288.0  dissatisfied
25904                     169.0     satisfied
25942                     138.0  dissatisfied

[505 rows x 25 columns]
```

[47]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import OneHotEncoder, StandardScaler
```

[63]:
```
df_new.replace({'satisfaction':{'satisfied':1,'dissatisfied':0}},inplace=True)
```

```
C:\Users\prera\AppData\Local\Temp\ipykernel_12352\3861770801.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_new.replace({'satisfaction':{'satisfied':1,'dissatisfied':0}},inplace=True)
```

[64]:
```
categorical_cols=['Gender','Customer Type','Type of Travel','Class']
encoder=OneHotEncoder(drop='first',sparse=False)
encoder_cols=pd.DataFrame(encoder.
 ↪fit_transform(df_new[categorical_cols]),columns=encoder.
 ↪get_feature_names_out(categorical_cols))
```

```
C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\preprocessing\_encoders.py:868: FutureWarning: `sparse` was
renamed to `sparse_output` in version 1.2 and will be removed in 1.4.
`sparse_output` is ignored unless you leave `sparse` to its default value.
  warnings.warn(
```

[65]:
```
numerical_cols=['id','Age','Flight Distance','Inflight wifi␣
 ↪service','Departure_Arrival time convenient','Ease of Online booking','Gate␣
 ↪location','Food and drink',
                'Online boarding','Seat comfort','Inflight␣
 ↪entertainment','Onboard service','Leg room service','Baggage handling',
                'Checkin service','Inflight service','Cleanliness','Departure␣
 ↪Delay in Minutes','Arrival Delay in Minutes']
scaler=StandardScaler()
```

```
scaled_cols=pd.DataFrame(scaler.
  ↪fit_transform(df_new[numerical_cols]),columns=scaler.
  ↪get_feature_names_out(numerical_cols))
```

[66]:
```
X=pd.concat([encoder_cols,scaled_cols],axis=1)
Y=df_new['satisfaction']
```

[67]:
```
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.
  ↪2,random_state=42)
log=LogisticRegression()
log.fit(X_train,Y_train)
```

[67]:
```
LogisticRegression()
```

[68]:
```
print("Train score",log.score(X_train,Y_train))
print("Test score",log.score(X_test,Y_test))
```

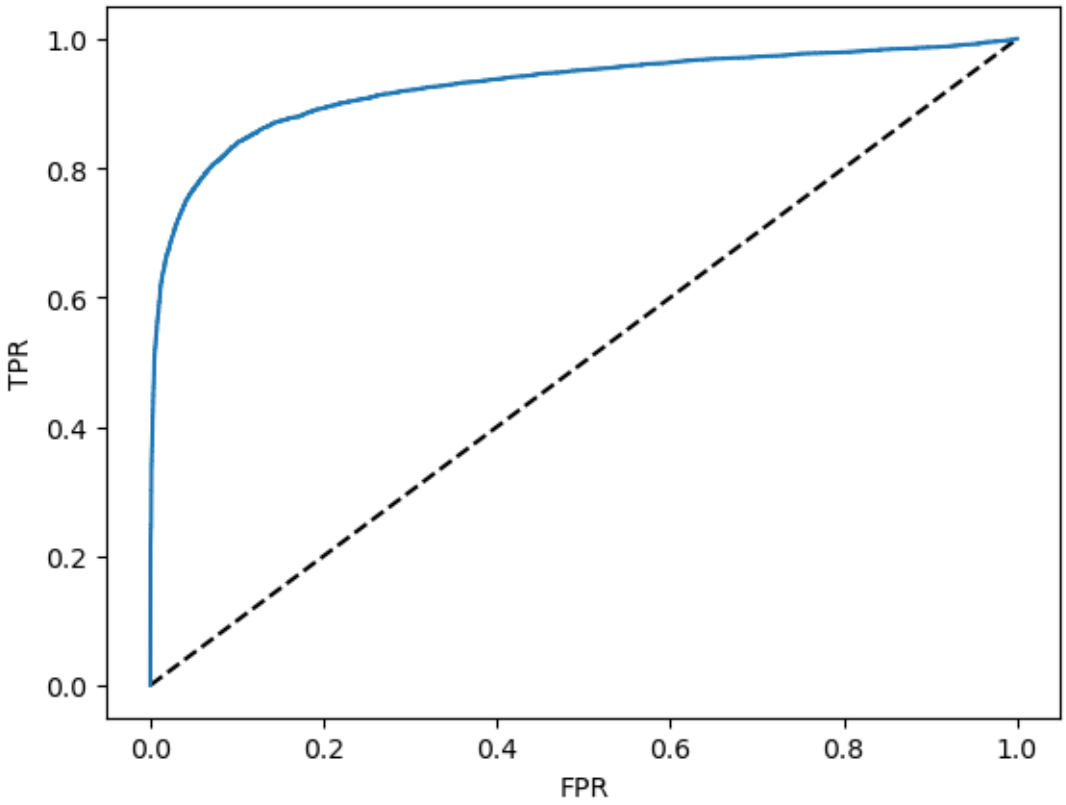
```
Train score 0.8730395076434385
Test score 0.8727669710202461
```

[69]:
```
pred_train = log.predict(X_train)
pred_test = log.predict(X_test)
```

[70]:
```
from sklearn import metrics
print(metrics.classification_report(Y_train,pred_train))
```

```
              precision    recall  f1-score   support

           0       0.88      0.90      0.89     11296
           1       0.87      0.84      0.85      8852

    accuracy                           0.87     20148
   macro avg       0.87      0.87      0.87     20148
weighted avg       0.87      0.87      0.87     20148
```

[72]:
```
roc=log.predict_proba(X_train)[:,1]
fpr, tpr, threshold = metrics.roc_curve(Y_train,roc)
plt.plot([0,1], [0,1],'k--') #x=(0,1) y
plt.plot(fpr,tpr,label='logistic')
plt.ylabel("TPR")
plt.xlabel("FPR")
plt.show()
```



[73]:
```
metrics.roc_auc_score(Y_train,roc)
```

[73]: 0.9263173768607853

[74]:
```
from sklearn.metrics import matthews_corrcoef

mcc= matthews_corrcoef(Y_test,pred_test)
print("MCC: ",mcc)
```

```
MCC:  0.7419059292213896
```

[75]:
```
from sklearn.model_selection import GridSearchCV
param_grid ={
    'penalty':["l1","l2"],
    'C': [0.1,0.5,1,5,10]
} #l1 lasso l2 ridge to regularize the model , C common factor alpha LOS
grid=GridSearchCV(estimator=log, param_grid=param_grid, cv=5)
grid.fit(X_train,Y_train)
```

```
C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\model_selection\_validation.py:378: FitFailedWarning:
25 fits failed out of a total of 50.
```

```
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
25 fits failed with the following error:
Traceback (most recent call last):
  File "C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\model_selection\_validation.py", line 686, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\linear_model\_logistic.py", line 1162, in fit
    solver = _check_solver(self.solver, self.penalty, self.dual)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\linear_model\_logistic.py", line 54, in _check_solver
    raise ValueError(
ValueError: Solver lbfgs supports only 'l2' or 'none' penalties, got l1 penalty.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\prera\anaconda3\Lib\site-
packages\sklearn\model_selection\_search.py:952: UserWarning: One or more of the
test scores are non-finite: [       nan 0.87219538        nan 0.87244353
nan 0.87239392
        nan 0.87219539        nan 0.87224502]
  warnings.warn(
```

[75]:
```
GridSearchCV(cv=5, estimator=LogisticRegression(),
             param_grid={'C': [0.1, 0.5, 1, 5, 10], 'penalty': ['l1', 'l2']})
```

[76]:
```
best_param = grid.best_params_
best_model = grid.best_estimator_
```

[77]:
```
y_pred=best_model.predict(X_test)
```

[79]:
```
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
 ↪f1_score, roc_auc_score
acc=accuracy_score(Y_test, y_pred)
pre=precision_score(Y_test, y_pred)
rec=recall_score(Y_test, y_pred)
f1=f1_score(Y_test, y_pred)
roc_auc=roc_auc_score(Y_test, y_pred)
print('Best param: ',best_param)
print('Accuracy: ', acc)
print('Precision: ',pre)
```

```
print('Recall: ',rec)
print('F1 Score: ',f1)
print('AUC-ROC: ',roc_auc)
```

```
Best param:  {'C': 0.5, 'penalty': 'l2'}
Accuracy:  0.8731639539499801
Precision:  0.8737185461323392
Recall:  0.8359340169415961
F1 Score:  0.8544087491455913
AUC-ROC:  0.869487580921603
```

[80]:
```
from sklearn.utils.fixes import loguniform
from sklearn.model_selection import RandomizedSearchCV
```

[83]:
```
logistic_random_param={'C':loguniform(1e-4,1e0),
                       'max_iter':(np.arange(100,800,10))}
grid=RandomizedSearchCV(estimator=log,␣
 ↪param_distributions=logistic_random_param, cv=5)
grid.fit(X_train,Y_train)
```

[83]:
```
RandomizedSearchCV(cv=5, estimator=LogisticRegression(),
                   param_distributions={'C':
<scipy.stats._distn_infrastructure.rv_continuous_frozen object at
0x00000160DB59F550>,
                                        'max_iter': array([100, 110, 120, 130,
140, 150, 160, 170, 180, 190, 200, 210, 220,
       230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350,
       360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480,
       490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610,
       620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740,
       750, 760, 770, 780, 790])})
```

[84]:
```
best_param = grid.best_params_
best_model = grid.best_estimator_
y_pred=best_model.predict(X_test)
```

[85]:
```
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
 ↪f1_score, roc_auc_score
acc=accuracy_score(Y_test, y_pred)
pre=precision_score(Y_test, y_pred)
rec=recall_score(Y_test, y_pred)
f1=f1_score(Y_test, y_pred)
roc_auc=roc_auc_score(Y_test, y_pred)
print('Best param: ',best_param)
print('Accuracy: ', acc)
print('Precision: ',pre)
print('Recall: ',rec)
```

```
print('F1 Score: ',f1)
print('AUC-ROC: ',roc_auc)
```

```
Best param:  {'C': 0.11747533383792395, 'max_iter': 700}
Accuracy:  0.8733624454148472
Precision:  0.8748248482017749
Recall:  0.8350423539901917
F1 Score:  0.854470802919708
AUC-ROC:  0.8695784220755967
```

[ ]: